

mini Manuel

de

Mathématiques pour la gestion

Benjamin Legros

- Statistiques
- Probabilités

**COURS
+
EXOS**

DUNOD

Le pictogramme qui figure ci-contre mérite une explication. Son objet est d'alerter le lecteur sur la menace que représente pour l'avenir de l'écrit, particulièrement dans le domaine de l'édition technique et universitaire, le développement massif du photocopillage.

Le Code de la propriété intellectuelle du 1^{er} juillet 1992 interdit en effet expressément la photocopie à usage collectif sans autorisation des ayants droit. Or, cette pratique s'est généralisée dans les établissements

d'enseignement supérieur, provoquant une baisse brutale des achats de livres et de revues, au point que la possibilité même pour

les auteurs de créer des œuvres nouvelles et de les faire éditer correctement est aujourd'hui menacée. Nous rappelons donc que toute reproduction, partielle ou totale, de la présente publication est interdite sans autorisation de l'auteur, de son éditeur ou du Centre français d'exploitation du droit de copie (CFC, 20, rue des Grands-Augustins, 75006 Paris).



© Dunod, Paris, 2011
ISBN 978-2-10-0559011-5

Le Code de la propriété intellectuelle n'autorisant, aux termes de l'article L. 122-5, 2° et 3° a), d'une part, que les « copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective » et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, « toute représentation ou reproduction intégrale ou partielle faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite » (art. L. 122-4).

Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles L. 335-2 et suivants du Code de la propriété intellectuelle.

Table des matières

Partie 1 Statistiques

1	Statistiques à une variable	3
	1.1 Étude statistique et représentation	3
	1.2 Indicateurs de tendance centrale	8
	1.3 Indicateurs de dispersion	12
	1.4 Interprétation des résultats	16
	Points clefs	17
	Exercices	18
	Solutions	19
2	Statistiques à deux variables	25
	2.1 Covariance	26
	2.2 Régression linéaire	29
	2.3 Régressions non linéaires	34
	Points clefs	36
	Exercices	37
	Solutions	39
3	Séries chronologiques	45
	3.1 Moyennes mobiles	47
	3.2 Résistance et support	49

3.3	Coefficients saisonniers	50
3.4	Phénomène de retracement	54
	Points clefs	54
	Exercices	55
	Solutions	57
4	Indices	65
4.1	Mesure d'évolution	65
4.2	Mode de calcul des indices	67
4.3	Exemples d'indices	72
	Points clefs	72
	Exercices	73
	Solutions	75
Partie 2		
Probabilités		
5	Notions de base de probabilités	79
5.1	Dénombrement	80
5.2	Calcul de probabilités	84
5.3	Lois de probabilités	89
	Points clefs	95
	Exercices	97
	Solutions	100
6	Lois fondamentales de probabilités	109
6.1	Loi binomiale	110
6.2	Loi de Poisson	115
6.3	Loi normale	117
6.4	Approximations de lois	122
	Points clefs	127
	Exercices	128
	Solutions	131

7	Estimateurs et tests d'hypothèses	141
	7.1 Échantillons	141
	7.2 Estimation d'une moyenne	143
	7.3 Estimation de proportion	146
	7.4 Différences de moyennes ou de proportions	147
	7.5 Tests d'hypothèses	148
	Points clefs	151
	Exercices	152
	Solutions	155
8	Test du χ^2	161
	8.1 Modélisation par une loi	162
	8.2 Test d'indépendance	164
	Points clefs	166
	Exercices	167
	Solutions	168
	Étude de cas Problème de synthèse	171

Comment utiliser le Mini Manuel ?

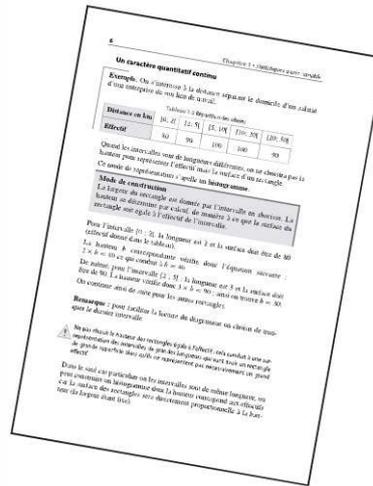
La page d'entrée de chapitre



Elle donne le plan du cours, ainsi qu'un rappel des objectifs pédagogiques du chapitre.

Le cours

Le cours, concis et structuré, expose les notions importantes du programme.



Les rubriques



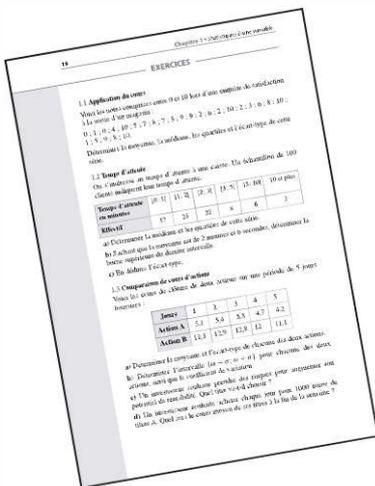
Une erreur à éviter



Un peu de méthode



Les points clés à retenir



Les exercices

Ils sont proposés en fin de chapitre, avec leur solution, pour se tester tout au long de l'année.

Chapitre 1	Statistiques à une variable	3
Chapitre 2	Statistiques à deux variables	25
Chapitre 3	Séries chronologiques	45
Chapitre 4	Indices	65

Prenons l'exemple des ventes d'un magasin, quels paramètres permettent-ils de comprendre un résultat annuel ? Il y a des facteurs externes : la situation économique, le niveau de vie des clients ou la présence de concurrents ; et des facteurs internes : la publicité, les prix ou la qualité de service. Cette liste est loin d'être exhaustive et l'influence réelle de chacun des paramètres est difficile à maîtriser. Pourtant, il est fondamental de comprendre au mieux les éléments essentiels influant sur ces ventes.

En économie et en finance, les problèmes ont en commun la multiplicité des éléments influant. L'outil des statistiques permet d'appréhender ce type de problèmes et de réaliser une synthèse des grandes dynamiques en présence.

La première partie de l'ouvrage présente les outils statistiques les plus utiles pour le gestionnaire.

Le premier chapitre, « Statistiques à une variable », sert à construire les outils d'observation d'un phénomène. Le second chapitre, « Statistiques à deux variables », permet d'apprécier le lien entre deux grandeurs et de l'utiliser à des fins de prévisions. Le troisième chapitre, « Séries chronologiques », est particulièrement utile à la finance de marché et dans les phénomènes variant dans le temps. Le quatrième chapitre, « Indices », présente un outil de compréhension des phénomènes économiques.

Statistiques à une variable

OBJECTIFS

- ▶ Savoir construire un diagramme adapté à une série statistique
- ▶ Maîtriser les paramètres de base de l'évaluation statistique (moyenne, médiane, écart-type et quartiles)
- ▶ Savoir interpréter un résultat à l'aide d'intervalles représentatifs ou du coefficient de variation

PLAN

- 1.1 Étude statistique et représentation
- 1.2 Indicateurs de tendance centrale
- 1.3 Indicateurs de dispersion
- 1.4 Interprétation des résultats

1.1 ÉTUDE STATISTIQUE ET REPRÉSENTATION

a) Vocabulaire de l'étude

L'analyse statistique consiste à extraire une information utile et synthétique d'un ensemble d'observations. L'étude doit se limiter à une **population** qui, pour le gestionnaire, sera par exemple la production d'une usine, les salariés d'une entreprise ou encore un ensemble de consommateurs d'un produit. L'analyse de cette population se fait au travers d'une grille de critères encore appelés **caractères**.

Ces caractères peuvent être **qualitatifs**. On entend par qualitatif un caractère qui ne peut pas être évalué par un chiffre ; par exemple, la couleur d'une voiture, le poste dans une entreprise ou la nationalité.

Ces caractères peuvent être **quantitatifs**. Selon le besoin de l'étude, on peut choisir de considérer chaque résultat individuellement ; on parle alors de **caractère discret**, par exemple, le résultat à une épreuve ou le nombre de voitures dont dispose un individu. Si le détail des résultats n'apporte pas de grand intérêt, on regroupe les résultats par intervalles. On parle alors de **caractères continus** ; par exemple, la distance du domicile d'un salarié à son lieu de travail. Il sera plus intéressant de savoir combien habitent entre 5 km et 10 km du lieu de travail plutôt que combien habitent précisément à 7 km.

b) La fréquence

Les statistiques sont un vecteur majeur de communication. L'actualité des entreprises regorge de données statistiques diverses qui par leurs présentations informent (ou désinforment).

Le premier élément de communication est la **fréquence** : on exprime l'importance d'une donnée sous forme de pourcentage.

Exemple. L'usine A a produit 457 objets dont 34 sont défectueux et l'usine B a produit 537 objets dont 42 sont défectueux.

Il n'est pas évident à première vue de dire quelle usine a la meilleure qualité de production. Si on exprime les choses ainsi : l'usine A a 7,44 % de produits défectueux et l'usine B en a 7,82 %, la comparaison est facile.

Pour cela, on utilise la **formule de la fréquence** :

$$\text{Fréquence} = \frac{\text{Effectif considéré}}{\text{Effectif Total}} \times 100$$

Le résultat est ainsi exprimé sous forme de pourcentage.

Exemple. Pour l'usine A, l'effectif considéré est 34, l'effectif total est 457. Avec la formule précédente on retrouve donc 7,44 %.

c) Modes de représentation

Comment représenter au mieux une série statistique ? La réponse dépend du type de la série étudiée ainsi que de l'information que l'on souhaite rendre visible. Il n'y a pas de règles absolues pour représenter une série : dans la littérature, on constate l'utilisation de tous types de diagrammes pour tous types de séries. Cependant, pour éviter d'induire de fausses informations, certains diagrammes semblent plus adaptés que d'autres.

Un caractère quantitatif discret

Exemple. Une population de 100 clients évaluent un centre d'appels téléphoniques par une note de 0 à 5 :

Tableau 1-1 Résultat de l'évaluation

Notes	0	1	2	3	4	5
Effectif	10	15	10	35	25	5

Il s'agit dans cet exemple de représenter la gradation des notes – 0 moins bon que 1 lui même moins bon que 2 ... – et l'importance de la représentation de chaque note donnée par l'effectif.

Pour figurer une série discrète, le mode de représentation le plus simple à produire et à interpréter est le **diagramme en bâton**.

Mode de construction

Dans un diagramme en bâtons, on représente les notes par des bâtons dont la hauteur est égale à l'effectif.

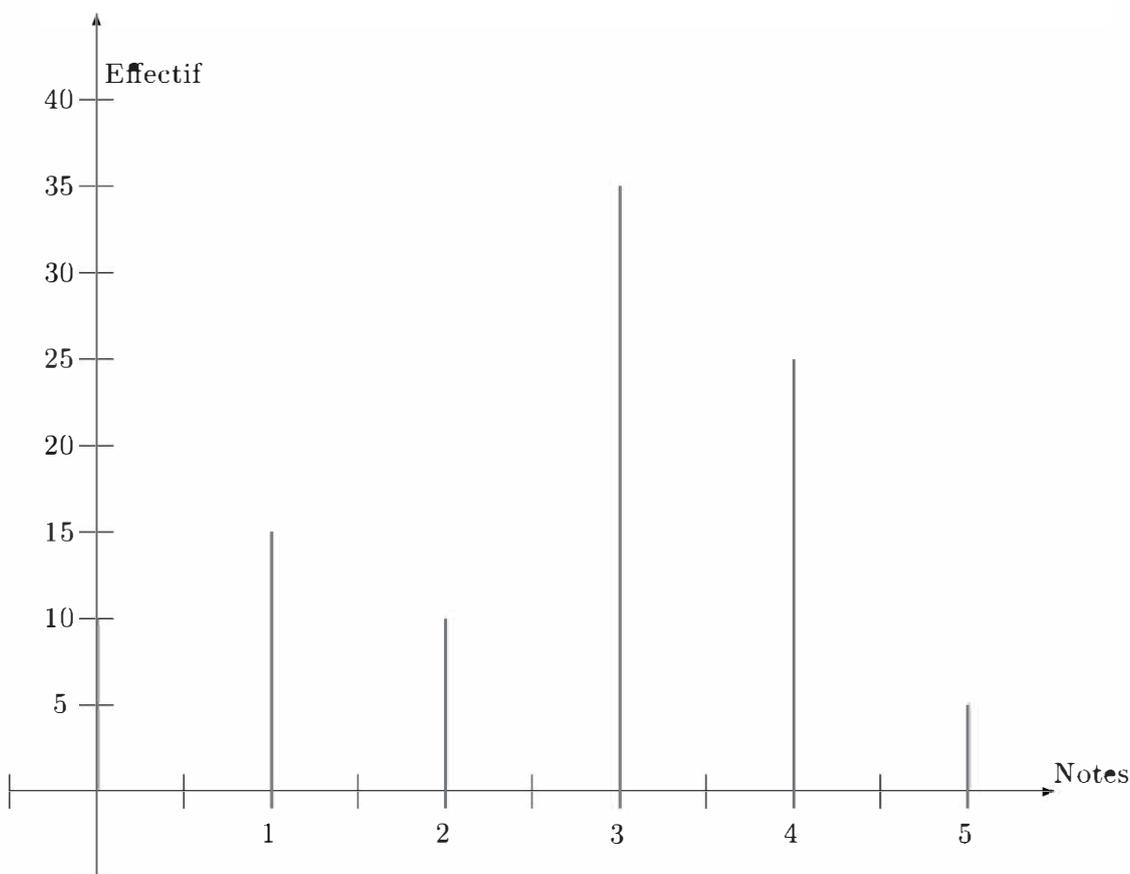


Figure 1-1 Diagramme en bâton

Un caractère quantitatif continu

Exemple. On s'intéresse à la distance séparant le domicile d'un salarié d'une entreprise de son lieu de travail.

Tableau 1-2 Répartition des salariés

Distance en km	[0; 2[[2; 5[[5; 10[[10; 20[[20; 50[
Effectif	80	90	100	100	90

Quand les intervalles sont de longueurs différentes, on ne choisira pas la hauteur pour représenter l'effectif mais la surface d'un rectangle.

Ce mode de représentation s'appelle un **histogramme**.

Mode de construction

La largeur du rectangle est donnée par l'intervalle en abscisse. La hauteur se détermine par calcul, de manière à ce que la surface du rectangle soit égale à l'effectif de l'intervalle.

Pour l'intervalle [0 ; 2[, la longueur est 2 et la surface doit être de 80 (effectif donné dans le tableau).

La hauteur h correspondante vérifie donc l'équation suivante : $2 \times h = 80$ ce qui conduit à $h = 40$

De même, pour l'intervalle [2 ; 5[; la longueur est 3 et la surface doit être de 90. La hauteur vérifie donc $3 \times h = 90$; ainsi on trouve $h = 30$.

On continue ainsi de suite pour les autres rectangles.

Remarque : pour faciliter la lecture du diagramme on choisit de tronquer le dernier intervalle.



Ne pas choisir la hauteur des rectangles égale à l'effectif ; cela conduit à une sur-représentation des intervalles de grandes longueurs qui vont avoir un rectangle de grande superficie alors qu'ils ne représentent pas nécessairement un grand effectif.

Dans le seul cas particulier où les intervalles sont de même longueur, on peut construire un histogramme dont la hauteur correspond aux effectifs car la surface des rectangles sera directement proportionnelle à la hauteur (la largeur étant fixe).

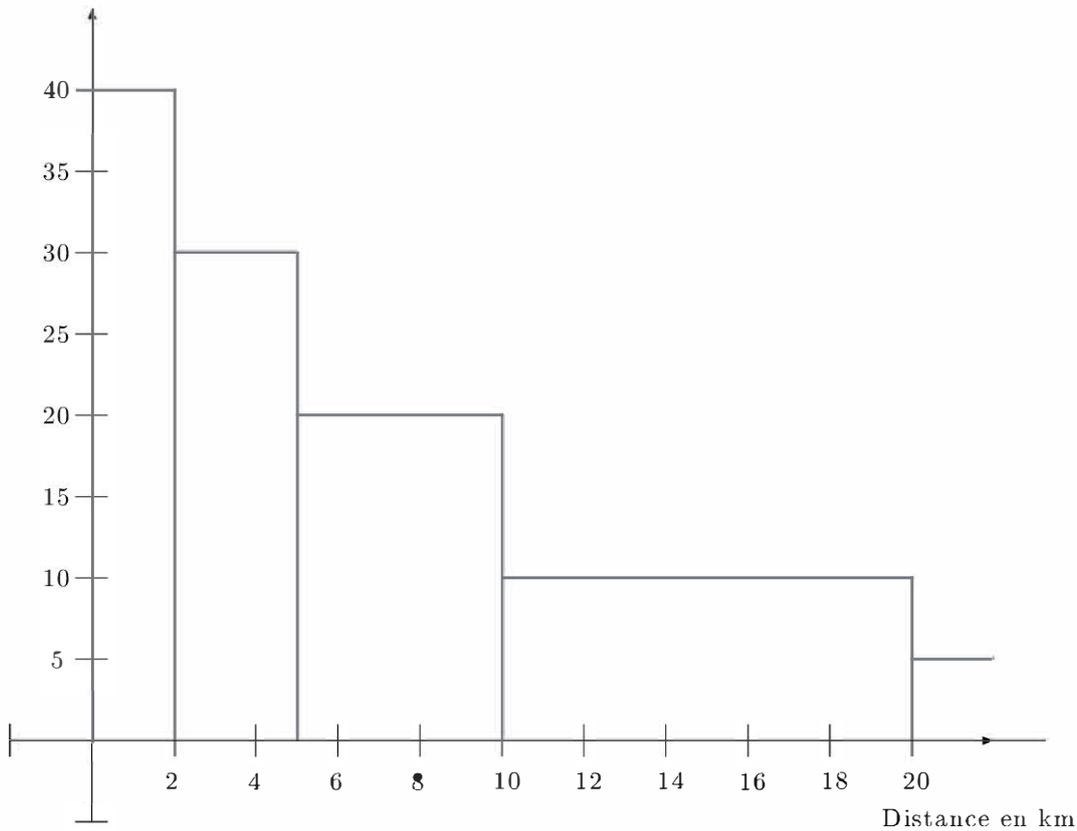


Figure 1-2 Histogramme

Un caractère qualitatif

Exemple. On s'intéresse aux couleurs des voitures d'une sortie de production pendant une période donnée. Voici les résultats constatés :

Tableau 1-3 Sortie de production

Couleurs	Noir	Rouge	Jaune	Vert
Production	10 000	5 000	2 000	3 000

L'utilisation d'un histogramme ou d'un diagramme en bâton induirait, dans le cas de données qualitatives, l'idée d'une progression du type « telle couleur meilleure que telle autre ». Ce qui n'est pas l'intention ici.

Il est plus judicieux de représenter cette série par un **diagramme circulaire** où l'angle considéré est proportionnel à l'effectif.

La forme circulaire évite de donner l'illusion d'une gradation. Pour retrouver un sens de lecture on utilise un diagramme **semi-circulaire**. L'exemple le plus classique est la composition politique d'un parlement.

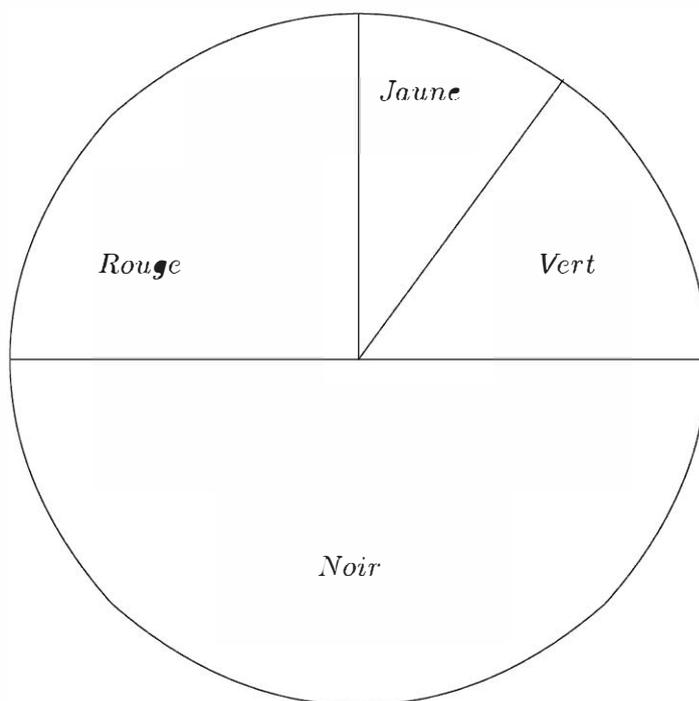


Figure 1-3 Diagramme circulaire

1.2 INDICATEURS DE TENDANCE CENTRALE

a) Moyennes

Une moyenne est une valeur qui se trouve au milieu de toutes les autres. C'est un indicateur de **tendance centrale** qui permet d'appréhender une population de manière globale. La moyenne, comme les autres éléments du calcul statistique, ne s'utilise que pour des caractères quantitatifs. On pourrait essayer de calculer la couleur moyenne des voitures de l'exemple précédent, mais cela ne serait d'aucune utilité.

Moyenne arithmétique simple

Exemple. Dans l'exemple suivant, on s'intéresse à la production mensuelle d'une usine sur une période de 6 mois :

Tableau 1-4 Production mensuelle

Mois	Janvier	Février	Mars	Avril	Mai	Juin
Production en unités	10 000	50 000	20 000	30 000	30 000	40 000

Chaque mois a la même importance ; si on cherche à calculer une moyenne de la production mensuelle, on ajoute les différentes productions pour les diviser par le nombre de mois :

$$\begin{aligned} \text{Moyenne} &= \frac{10\,000 + 50\,000 + 20\,000 + 30\,000 + 30\,000 + 40\,000}{6} \\ &= 30\,000 \end{aligned}$$

En notant x_1, x_2, \dots, x_n les valeurs de la série et \bar{x} la moyenne, on a la formule générale d'une **moyenne simple** :

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

Moyenne arithmétique pondérée

Dans le cas du *Tableau 1-1*, on va chercher à déterminer une note moyenne de satisfaction. Les effectifs associés à chaque note sont à prendre en considération.

En notant e_1, e_2, \dots, e_n les effectifs associés aux valeurs x_1, x_2, \dots, x_n et \bar{x} la moyenne, on a la formule générale d'une **moyenne pondérée** :

$$\bar{x} = \frac{e_1 x_1 + e_2 x_2 + \dots + e_n x_n}{e_1 + e_2 + \dots + e_n} = \frac{\sum_{i=1}^n e_i x_i}{\sum_{i=1}^n e_i}$$

Ce qui, appliqué à la série du *Tableau 1-1*, conduit à :

$$\begin{aligned} \text{Moyenne} &= \frac{0 \times 10 + 1 \times 15 + 2 \times 10 + 3 \times 35 + 4 \times 25 + 5 \times 5}{100} \\ &= 2,65 \end{aligned}$$

Moyenne pour un caractère continu

Il se pose le problème du choix de la valeur à considérer pour appliquer la formule précédente. La solution la plus pratique est de retenir les milieux de chacune des classes comme valeurs. Dans l'exemple du *Tableau 1-2*, on trouve :

$$\begin{aligned} \text{Moyenne} &= \frac{1 \times 80 + 3,5 \times 90 + 7,5 \times 100 + 15 \times 100 + 35 \times 90}{80 + 90 + 100 + 100 + 90} \\ &\approx 12,60 \text{ km} \end{aligned}$$

Remarque : cette valeur est en réalité le résultat d'une approximation ; on a supposé que les effectifs étaient répartis de manière homogène dans les intervalles. Cela justifie le choix du milieu des intervalles. On procédera de même pour les calculs d'écart-moyen et d'écart-type.

b) Médiane

On considère la suite de notes suivantes : 7 ; 8 ; 9 ; 20. La moyenne de cette série est 11. Il s'agit bien du milieu de ces 4 notes ; pour autant 75 % de ces notes sont en dessous de 10 et la moyenne semble donner une indication positive à savoir 11. La moyenne est en effet très influencée par les valeurs extrêmes, ici le 20.

Il est donc utile de s'intéresser à un autre indicateur de tendance centrale qui est **la médiane**.

Définition : la médiane est la valeur de la série pour laquelle 50 % de la population a ses valeurs en dessous et 50 % a ses valeurs au dessus.

Ici, la médiane est entre 8 et 9, elle est de 8,5. Cette valeur de 8,5 indique bien le fait que 3 notes sur 4 sont en dessous de 10 et que l'évaluation n'est pas aussi positive que semblait l'indiquer la moyenne.

Un autre exemple de différence sensible entre moyenne et médiane est le salaire. On note un salaire moyen net en France autour de 1 800 € et un salaire médian net autour de 1 500 €. L'explication est la même que pour la série précédente ; des salaires élevés mais peu nombreux augmentent la moyenne des salaires mais n'ont que peu d'effet sur la médiane.

Calcul pour un caractère discret

Commençons par des exemples simples :

Exemple.

Série a : - 1 ; 3 ; 6 ; 8 ; 9

Les valeurs sont classées par ordre croissant, il y a 5 valeurs. La valeur qui sépare la série en deux sous-séries de même taille est 6. Ainsi la médiane est 6.

Exemple.

Série b : 4 ; 24 ; 32 ; 50

Il y a ici 4 valeurs dans l'ordre croissant. Aucune de ces valeurs ne sépare la série en deux. La médiane se situe entre les deux valeurs du milieu ; entre 24 et 32. La médiane est donc 28.



Méthode générale pour le calcul de la médiane

On note N le nombre de valeurs de la série.

1. On classe les valeurs par ordre croissant.

2. Deux cas se présentent :

– soit N est impair et on calcule $\frac{N+1}{2}$, la médiane est alors la valeur située à la position $\frac{N+1}{2}$

– soit N est pair et on calcule $\frac{N}{2}$, la médiane est alors la valeur située entre la valeur à la position $\frac{N}{2}$ et la valeur à la position $\frac{N}{2} + 1$.

Application au cas du *Tableau 1-1* : on a $N = 100$, donc la médiane est entre la 50^e et la 51^e valeur. On établit le tableau des effectifs cumulés croissants :

Notes	0	1	2	3	4	5
Effectif	10	15	10	35	25	5
Effectifs cumulés croissants	10	25	35	70	95	100

On constate ainsi que la 50^e valeur est 3 et la 51^e valeur est aussi 3. Ainsi la médiane est 3.

Calcul pour un caractère continu

Reprenons l'exemple du *Tableau 1-2*. L'effectif total est $80 + 90 + 100 + 100 + 90 = 460$. On calcule $\frac{460}{2} = 230$. Ainsi, la médiane se situe exactement entre la valeur de la 230^e et la valeur de la 231^e personne.

Complétons le tableau avec les effectifs cumulés croissants :

Distance en km	[0; 2[[2; 5[[5; 10[[10; 20[[20; 50[
Effectif	80	90	100	100	90
Effectifs cumulés croissants	80	170	270	370	460

La dernière ligne du tableau indique que 170 personnes vivent entre 0 et 5 km de leur lieu de travail et 270 personnes vivent entre 0 et 10 km de leur lieu de travail. La médiane est donc dans l'intervalle [5 ; 10[et correspond à un effectif cumulé de 230.

On peut représenter schématiquement la position de la médiane M ainsi :

$$\begin{array}{ccc} 5 & M & 10 \\ | & | & | \\ \hline 170 & 230 & 270 \end{array}$$

En supposant une répartition parfaitement homogène de la population dans chaque intervalle, on a :

$$\frac{M - 5}{230 - 170} = \frac{10 - 5}{270 - 170}$$

$$\text{ainsi : } \frac{M - 5}{60} = \frac{5}{100} \text{ et } M - 5 = \frac{60 \times 5}{100} \text{ d'où } M = 5 + 3 = 8 \text{ km}$$

Cette méthode est reproductible. On retrouvera ce résultat à l'aide du diagramme des effectifs cumulés.

c) Mode

Dans le cas d'une série discrète, le **mode** est la valeur ayant le plus grand effectif. Dans l'exemple du *Tableau 1-1*, il s'agit de 3.

Dans le cas d'une série continue, on parle de **classe modale** pour évoquer la classe ayant le plus grand effectif ; le **mode** est le milieu de cette classe.

La notion de mode n'a d'intérêt que pour les séries pour lesquelles une valeur se détache des autres par l'importance de son effectif. Dans le cas du *Tableau 1-2*, la recherche du mode ne présente pas d'intérêt, l'ensemble des classes ayant des effectifs proches.

1.3 INDICATEURS DE DISPERSION

Comparons les deux séries suivantes :

Série a : 0 ; 1 ; 19 ; 20

Série b : 8 ; 9 ; 11 ; 12

Ces deux séries ont la même moyenne et la même médiane, cependant elles sont significativement différentes. La différence essentielle entre elles est leur étalement ; la première série a des valeurs éloignées de la moyenne alors que la seconde a des valeurs proches de la moyenne. En statistiques, on utilise le terme de **dispersion** des valeurs. Il s'agit donc maintenant de présenter des outils de mesure de la dispersion d'une série.

a) Écart-moyen

L'écart-moyen est la mesure de dispersion la plus intuitive. Il se définit comme la moyenne des écarts à la moyenne de la série (les écarts étant toujours comptés positivement).

En notant m la moyenne de la série ; e_1, e_2, \dots, e_n les effectifs associés aux valeurs x_1, x_2, \dots, x_n , on a la formule suivante :

$$\begin{aligned} \text{Écart-moyen} &= \frac{e_1|x_1 - m| + e_2|x_2 - m| + \dots + e_n|x_n - m|}{e_1 + e_2 + \dots + e_n} \\ &= \frac{\sum_{i=1}^n e_i|x_i - m|}{\sum_{i=1}^n e_i} \end{aligned}$$

Notons que les valeurs absolues dans la formule servent à mesurer l'écart à la moyenne positivement.

Exemple du *Tableau 1-1*

La moyenne a pour valeur 2,65.

On trouve l'écart-moyen ainsi :

Écart-moyen

$$\begin{aligned} &= \frac{10 \times |0 - 2,65| + 15 \times |1 - 2,65| + 10 \times |2 - 2,65| + 35 \times |3 - 2,65| + 25 \times |4 - 2,65| + 5 \times |5 - 2,65|}{100} \\ &\approx 1,16 \end{aligned}$$

Notons que l'écart-moyen est très peu utilisé en comparaison de l'écart-type. L'écart-type a une grande importance en probabilité, au niveau de la loi normale en particulier, qui justifie son utilisation majoritaire pour mesurer la dispersion d'une série.

b) Écart-type

L'écart-type est l'élément le plus utilisé pour mesurer la dispersion d'une série.

Voici la formule définissant l'écart-type : notons σ l'écart-type, m la moyenne et e_1, e_2, \dots, e_n les effectifs associés aux valeurs x_1, x_2, \dots, x_n .

$$\sigma = \sqrt{\frac{e_1(x_1 - m)^2 + e_2(x_2 - m)^2 + \dots + e_n(x_n - m)^2}{e_1 + e_2 + \dots + e_n}}$$

$$= \sqrt{\frac{\sum_{i=1}^n e_i(x_i - m)^2}{\sum_{i=1}^n e_i}}$$

En développant cette formule, on peut établir la propriété suivante :

$$\text{Écart-type} = \sqrt{\text{Moyenne des carrés} - \text{Moyenne au carré}}$$

Voyons à partir du *Tableau 1-2*, comment utiliser cette formule pour calculer l'écart-type de la série.

1. Calcul de la moyenne des carrés

$$\frac{80 \times 1^2 + 90 \times 3,5^2 + 100 \times 7,5 + 100 \times 15^2 + 90 \times 35^2}{80 + 90 + 100 + 100 + 90} \approx 303,39$$

2. Calcul de la moyenne au carré

La moyenne a déjà été calculée au § 2.1.1

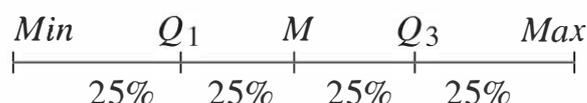
On a donc $12,6^2 = 158,76$

3. Calcul de l'écart-type

On trouve donc $\sigma = \sqrt{303,39 - 158,76} = 12,03$

c) Quartiles

Les quartiles sont des valeurs qui permettent de couper une population en quatre sous-populations de même taille. Le premier quartile correspond au premier quart de la population. On le note Q_1 . Le second quartile correspond au second quart de la population, il s'agit donc de la médiane. Le troisième quartile correspond au troisième quart. On le note Q_3 . Ils se répartissent schématiquement ainsi :



La méthode pour déterminer les quartiles est en tout point semblable à celle qui permet de déterminer la médiane.

d) Diagramme des effectifs cumulés

Le diagramme des effectifs cumulés sert à retrouver de manière graphique les paramètres de médiane et quartiles.

Mode de construction

À partir de la série du *Tableau 1-2* :

1. On complète le tableau avec la ligne effectifs cumulés croissants.

Distance en km	[0; 2[[2; 5[[5; 10[[10; 20[[20; 50[
Effectif	80	90	100	100	90
Effectifs cumulés croissants	80	170	270	370	460

2. On place les points dont les coordonnées sont les extrémités des intervalles avec l'effectif cumulé associé.

3. On relie les points par des segments de droite.

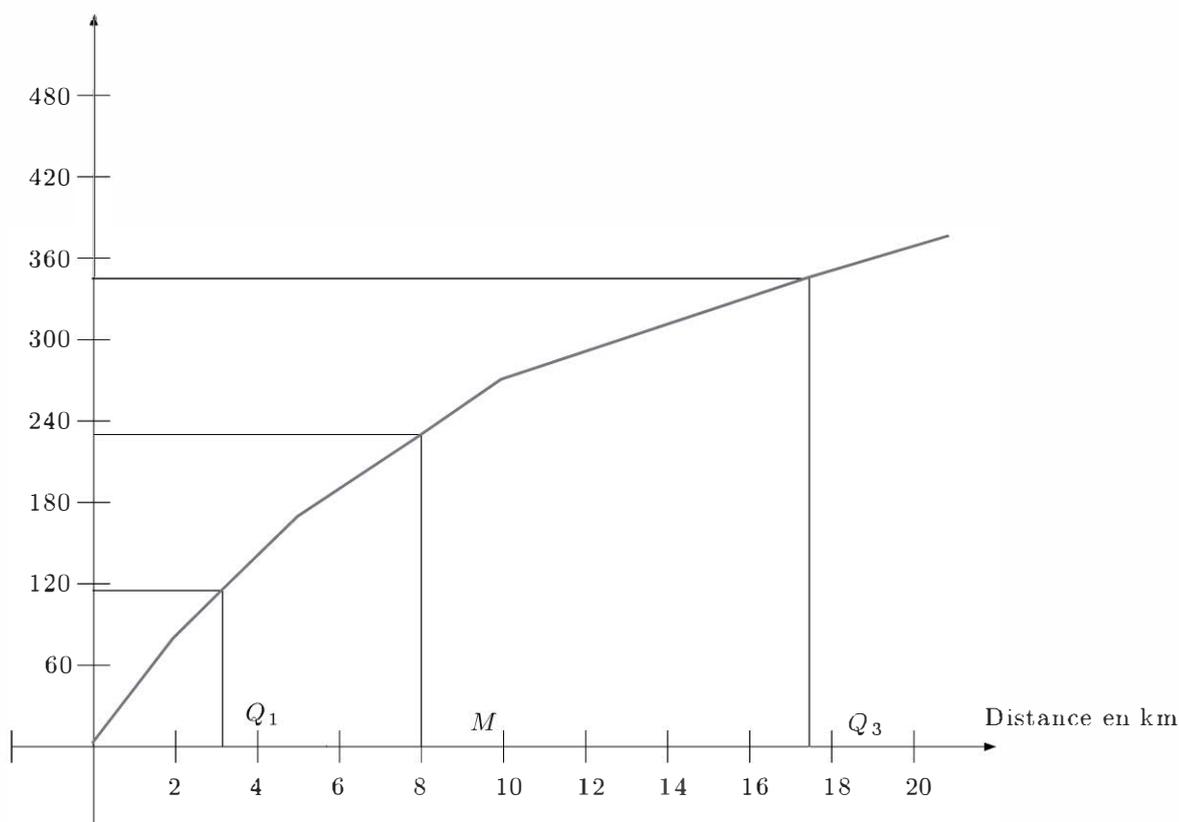


Figure 1-4 Diagramme des effectifs cumulés



Méthode pour retrouver la médiane et les quartiles

1. On calcule l'effectif cumulé correspondant à chacun de ces paramètres. Ici l'effectif total est 460, la médiane correspond donc à un effectif cumulé de $230 \left(\frac{460}{2} \right)$, le premier quartile de $115 \left(\frac{460}{4} \right)$ et le troisième de $345 \left(\frac{3 \times 460}{4} \right)$.

2. On place ces valeurs sur l'axe des ordonnées.

3. On trouve à l'aide du graphe l'abscisse correspondant à ces valeurs d'ordonnées ; ce qui permet de mesurer avec la précision du graphique : $Q_1 \approx 3$; $M \approx 8$ et $Q_3 \approx 17,5$

Par une méthode calculatoire, on trouve avec précision : $Q_1 = 3,17$; $M = 8$ et $Q_3 = 17,5$

Ainsi la méthode graphique permet avec moins de précision (mais aussi moins de calculs) de trouver les valeurs de Q_1 , Q_3 et M .

1.4 INTERPRÉTATION DES RÉSULTATS

a) Coefficient de variation

L'écart-moyen et l'écart-type mesurent la dispersion : ainsi, plus leur valeur est grande plus la série est étalée. Mais que signifie un écart-type grand ?

Rien dans l'absolu, un écart-type de 10 sera très grand pour une série dont les valeurs sont entre 0 et 30 et il sera très petit si les valeurs sont de l'ordre du million.

C'est pour cela qu'on utilise le coefficient de variation : $CV = \frac{\sigma}{m}$ où σ est l'écart-type et m est la moyenne. Ainsi, une valeur importante de CV indique une dispersion importante relativement à la moyenne.

b) Intervalles significatifs

- **L'intervalle interquartile** est l'intervalle $[Q_1; Q_3]$. Cet intervalle contient 50 % de la population. Il sert à déterminer la moitié centrale de la population.
- **L'intervalle de confiance** $[m - \sigma; m + \sigma]$ contient, dans une répartition normale, 68 % de la population. Il peut s'avérer utile de comparer la série à une distribution normale.



➤ **Mode de représentation d'une série statistique**

Type de série	Diagramme correspondant	Mode de construction
Caractère qualitatif	Diagramme circulaire	Angles proportionnels aux effectifs
Caractère quantitatif discret	Diagramme en bâton	Hauteurs proportionnelles aux effectifs
Caractère quantitatif continu	Histogramme	Surfaces proportionnelles aux effectifs

➤ **Calcul de moyenne**

- Prendre en compte la pondération par les effectifs lorsqu'il y a en a.
- Choisir le milieu des classes pour les séries à caractère continu.

➤ **Écart-type**

Pour une plus grande rapidité des calculs : calculer la moyenne des carrés, puis la moyenne au carré et prendre la racine carrée de la différence de ces deux valeurs.

➤ **Médiane et quartiles**

- Pour une série à caractère discret, différencier les cas selon la parité de l'effectif total.
- Pour une série à caractère continu, utiliser les effectifs cumulés pour poser une équation qui permet de trouver le résultat.

➤ **Interprétation des résultats**

- L'importance de la dispersion est relative à la moyenne, d'où l'importance du calcul du coefficient de variation.
- On peut s'intéresser à l'intervalle $[m - \sigma ; m + \sigma]$ dans lequel 68 % de la population se situe dans le cas d'une répartition normale, ou à l'intervalle interquartile $[Q_1 ; Q_3]$ dans lequel se situe 50 % de la population.

EXERCICES

1.1 Application du cours

Voici les notes comprises entre 0 et 10 lors d'une enquête de satisfaction à la sortie d'un magasin :

0 ; 1 ; 0 ; 4 ; 10 ; 5 ; 7 ; 8 ; 7 ; 5 ; 9 ; 9 ; 2 ; 6 ; 2 ; 10 ; 2 ; 3 ; 6 ; 8 ; 10 ; 1 ; 5 ; 9 ; 8 ; 10.

Déterminer la moyenne, la médiane, les quartiles et l'écart-type de cette série.

1.2 Temps d'attente

On s'intéresse au temps d'attente à une caisse. Un échantillon de 100 clients indiquent leur temps d'attente.

Temps d'attente en minutes	[0; 1[[1; 2[[2; 3[[3; 5[[5; 10[10 et plus
Effectif	37	25	22	8	6	2

- Déterminer la médiane et les quartiles de cette série.
- Sachant que la moyenne est de 2 minutes et 6 secondes, déterminer la borne supérieure du dernier intervalle.
- En déduire l'écart-type.

1.3 Comparaison de cours d'actions

Voici les cours de clôture de deux actions sur une période de 5 jours boursiers :

Jours	1	2	3	4	5
Action A	5,1	5,4	5,8	4,7	4,2
Action B	12,3	12,9	12,8	12	11,1

- Déterminer la moyenne et l'écart-type de chacune des deux actions.
- Déterminer l'intervalle $[m - \sigma; m + \sigma]$ pour chacune des deux actions, ainsi que le coefficient de variation.
- Un investisseur souhaite prendre des risques pour augmenter son potentiel de rentabilité. Quel titre va-t-il choisir ?
- Un investisseur souhaite acheter chaque jour pour 1000 euros de titres A. Quel sera le cours moyen de ses titres à la fin de la semaine ?

1.4 Normes d'un groupe industriel

Un groupe industriel fixe pour ses usines un objectif de production journalière moyenne de 1 000 unités avec un écart-type de 100 unités.

Pendant un mois de 30 jours, une usine a déterminé sa production :

Production en unités	[0; 200[[200; 500[[500; 1 000[[1 000; 1 500[[1 500; 2 000[
Nombre de jours	5	8	12	3	2

- Représenter graphiquement cette série.
- Etablir un diagnostic de la situation de l'usine en calculant la moyenne et l'écart-type de la production pendant le mois.
- Un manager propose une augmentation de la production journalière de 350 unités. Cette proposition est-elle satisfaisante ? Quelle proposition pourriez-vous formuler ?

SOLUTIONS

Exercice 1.1

La première étape est de regrouper les notes par ordres croissant et de déterminer les effectifs cumulés dans un tableau :

Notes	0	1	2	3	4	5	6	7	8	9	10
Effectif	2	2	3	1	1	3	2	2	3	3	4
Effectifs cumulés	2	4	7	8	9	12	14	16	19	22	26

La note moyenne

$$m = \frac{0 \times 2 + 1 \times 2 + 2 \times 3 + 3 \times 1 + 4 \times 1 + 5 \times 3 + 6 \times 2 + 7 \times 2 + 8 \times 3 + 9 \times 3 + 10 \times 4}{26}$$

$$\approx 5,65$$

L'écart-type

$$\sigma = \sqrt{\text{moyenne des carrés} - \text{moyenne au carré}}$$

Moyenne des carrés

$$= \frac{0^2 \times 2 + 1^2 \times 2 + 2^2 \times 3 + 3^2 \times 1 + 4^2 \times 1 + 5^2 \times 3 + 6^2 \times 2 + 7^2 \times 2 + 8^2 \times 3 + 9^2 \times 3 + 10^2 \times 4}{26}$$

$$\approx 43,04$$

$$\text{Moyenne au carré} = 5,65^2 \approx 31,97$$

Remarque : Il est important de ne pas faire des arrondis trop tôt dans les calculs. La valeur de 31,97 est calculée avec l'ensemble des chiffres en

mémoire dans la calculatrice pour la moyenne, à savoir 5,6538461538462 et non seulement 5,65.

$$\text{Ainsi } \sigma = \sqrt{43,04 - 31,97} \approx 3,33$$

La médiane

L'effectif total est de 26. La médiane M est donc entre la 13^e et la 14^e valeur. Le tableau indique que la 12^e valeur est un 5, la 13^e et la 14^e sont des 6. Ainsi $M = 6$.

Les quartiles

26 n'est pas divisible par 4. Considérons la première moitié de la série constituée des 13 premières valeurs. Le premier quartile Q_1 joue un rôle de médiane pour la première moitié de la série. On peut donc appliquer la méthode du cours. L'effectif de cette sous-série est impair. On calcule donc

$$\frac{13 + 1}{2} = 7. \text{ Ainsi, } Q_1 \text{ est exactement la 7^e valeur.}$$

La ligne des effectifs cumulés indique ainsi que $Q_1 = 2$.

Même raisonnement avec la seconde moitié de la série qui contient aussi 13 valeurs. Le troisième quartile est donc au niveau de la 7^e valeur de cette sous série donc au niveau de la valeur $13 + 7 = 20$.

La ligne des effectifs cumulés indique ainsi que $Q_3 = 9$.

Exercice 1.2

a) On ajoute au tableau une ligne pour les effectifs cumulés :

Temps d'attente en minutes	[0;1[[1;2[[2;3[[3;5[[5;10[10 et plus
Effectif	37	25	22	8	6	2
Effectifs cumulés	37	62	84	92	98	100

Éléments	Q_1	M	Q_3
Effectif cumulé	25	50	75
Schéma	$\begin{array}{c} 0 \qquad \qquad \qquad 1 \\ \qquad \qquad \qquad \\ \hline 0 \qquad \qquad \qquad 25 \qquad \qquad \qquad 37 \end{array}$	$\begin{array}{c} 1 \qquad \qquad \qquad 2 \\ \qquad \qquad \qquad \\ \hline 37 \qquad \qquad \qquad 50 \qquad \qquad \qquad 62 \end{array}$	$\begin{array}{c} 2 \qquad \qquad \qquad 3 \\ \qquad \qquad \qquad \\ \hline 62 \qquad \qquad \qquad 75 \qquad \qquad \qquad 84 \end{array}$
Équation	$\frac{Q_1 - 0}{25 - 0} = \frac{1 - 0}{37 - 0}$	$\frac{M - 1}{50 - 37} = \frac{2 - 1}{62 - 37}$	$\frac{Q_3 - 2}{75 - 62} = \frac{3 - 2}{84 - 62}$
Solution	$Q_1 = \frac{25}{37} \approx 0,67 \text{ min} = 41 \text{ s}$	$M = 1,52 \text{ min} = 1 \text{ min } 31 \text{ s}$	$Q_3 = 2,59 \text{ min} = 2 \text{ min } 35 \text{ s}$

b) Notons x la borne supérieure du dernier intervalle. Le milieu du dernier intervalle est donc $\frac{10+x}{2}$

La moyenne est de 2 minutes et 6 secondes, soit 2,10 minutes.

$$\text{Ainsi : } 2,10 = \frac{37 \times 0,5 + 25 \times 1,5 + 22 \times 2,5 + 8 \times 4 + 6 \times 7,5 + 2 \times \frac{10+x}{2}}{100}$$

$$\text{d'où } 2,10 = \frac{188 + 10 + x}{100} \text{ et } 210 = 198 + x, \text{ soit } x = 12 \text{ min.}$$

Le dernier intervalle est donc l'intervalle $[10 ; 12[$.

c) L'écart-type

On calcule la moyenne des carrés :

$$\frac{37 \times 0,5^2 + 25 \times 1,5^2 + 22 \times 2,5^2 + 8 \times 4^2 + 6 \times 7,5^2 + 2 \times 11^2}{100} = 9,105$$

$$\text{La moyenne au carré : } 2,1^2 = 4,41$$

Ainsi, on a $\sigma = \sqrt{9,105 - 4,41} = 2,16$ minutes = 2 minutes et 10 secondes.

Exercice 1.3

a) Pour la moyenne et l'écart-type, une erreur à ne pas commettre est de pondérer les cours par le jour. Cela n'a aucun sens, car cela reviendrait à donner une importance plus grande au dernier jour qu'au premier. Ce n'est pas l'intention ici.

On fait donc une **moyenne simple** :

$$\text{Pour A : } \frac{5,1 + 5,4 + 5,8 + 4,7 + 4,2}{5} = 5,04$$

$$\text{et pour B : } \frac{12,3 + 12,9 + 12,8 + 12 + 11,1}{5} = 12,22$$

On trouve l'écart-type de A :

$$\sigma_A = \sqrt{\frac{5,1^2 + 5,4^2 + 5,8^2 + 4,7^2 + 4,2^2}{5} - 5,04^2} \approx 0,553$$

et l'écart-type de B :

$$\sigma_B = \sqrt{\frac{12,3^2 + 12,9^2 + 12,8^2 + 12^2 + 11,1^2}{5} - 12,22^2} \approx 0,649$$

b) L'intervalle $[m - \sigma ; m + \sigma]$

Pour l'action A : $[5,04 - 0,55 ; 5,04 + 0,55] = [4,49 ; 5,59]$

Pour l'action B : $[12,22 - 0,65 ; 12,22 + 0,65] = [11,57 ; 12,87]$

Les coefficients de variation

$$CV_A = \frac{0,55}{5,04} = 10,91 \% \text{ et } CV_B = \frac{0,65}{12,22} = 5,32 \%$$

c) Un investisseur souhaite acheter le titre le plus rentable. Voyons l'utilité des différents paramètres calculés :

– **la moyenne** ; elle varie du simple au double entre les deux titres, mais cela n'a pas d'importance. En effet, un investisseur qui possède un certain capital investira le même montant quelle que soit la valeur des titres, il en achètera simplement moins si le titre a un cours plus élevé. Ce qui intéresse l'investisseur est le potentiel de variation du cours. **L'écart-type** semble donc intéressant ; l'action B a un écart-type légèrement plus grand que celui de l'action A. Cette différence n'est pas significative et doit être rapportée à la valeur de la moyenne.

Les intervalles sont intéressants à la condition d'avoir l'assurance d'une répartition gaussienne des cours des actions et une répartition globalement semblable d'une semaine à l'autre. En admettant cela, on peut estimer une valeur maximale de vente du titre de 5,59 euros pour l'action A et de 12,87 euros pour l'action B. Le pourcentage de progression dépend du cours auquel l'investisseur a acheté son titre.

Le coefficient de variation est l'élément le plus significatif ici. Exprimé en pourcentage, il indique le potentiel de hausse ou de baisse par rapport à la moyenne. Ainsi, l'action A a le plus fort potentiel de progrès. C'est donc ce titre que l'investisseur va choisir en sachant qu'il s'agit aussi du titre le plus risqué.

d) La question se pose souvent ainsi ; on ne fixe pas la quantité de titres à acheter, par contre on fixe le montant à investir, ici 1 000 euros par jour. On ne va pas faire une moyenne simple des cours ici, car le montant est le même mais le nombre d'actions achetées chaque jour varie. Déterminons tout d'abord le nombre d'actions achetées à la fin de la semaine. Une simple proportionnalité permet d'établir un achat de $\frac{1\,000}{5,1}$ la jour 1, $\frac{1\,000}{5,4}$ le jour 2 et ainsi de suite. Ainsi le nombre d'actions achetées en fin de semaine sera : $\frac{1\,000}{5,1} + \frac{1\,000}{5,4} + \frac{1\,000}{5,8} + \frac{1\,000}{4,7} + \frac{1\,000}{4,2}$. Le tout pour un montant total de 5 000 €. On cherche maintenant un coût unitaire d'achat des titres. Pour cela il suffit de faire le rapport suivant :

$\frac{1\,000}{5,1} + \frac{1\,000}{5,4} + \frac{5\,000}{5,8} + \frac{1\,000}{4,7} + \frac{1\,000}{4,2} \approx 4,977$. Un titre A a coûté en moyenne à l'investisseur 4,977 €.

Remarque : La moyenne ici n'est pas une moyenne arithmétique simple. Il s'agit d'une **moyenne harmonique**.

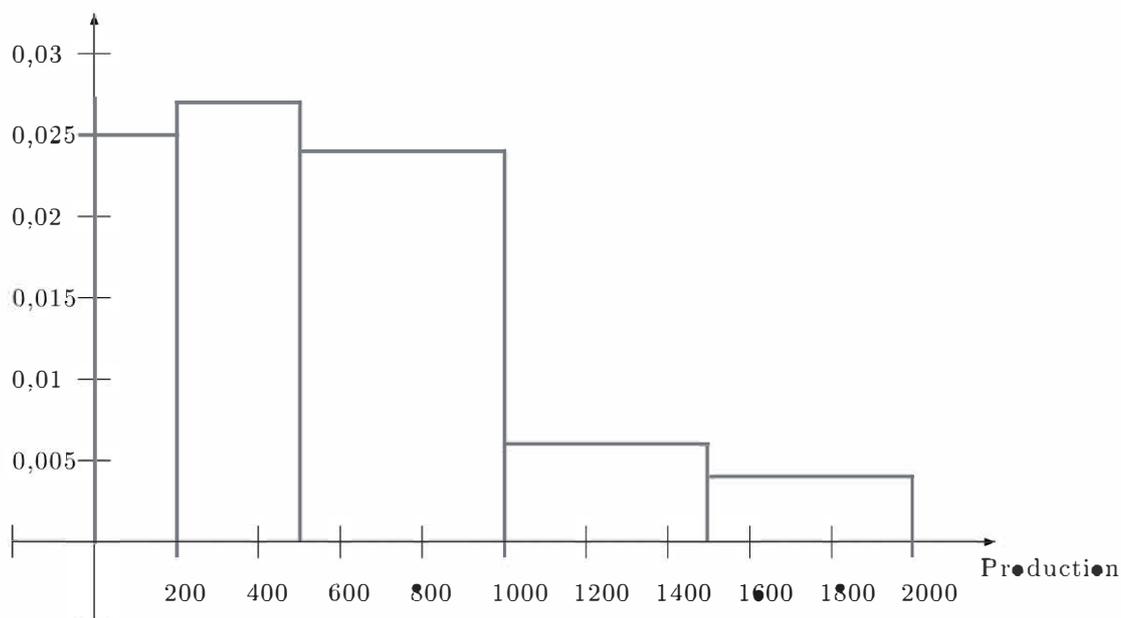
Exercice 1.4

a) On détermine les hauteurs des rectangles qui permettent de représenter les effectifs.

Voici les résultats sous forme de tableau :

Intervalle	Longueur de l'intervalle	Effectif	Hauteur
[0 ; 200[200	5	0,025
[200 ; 500[300	8	0,027
[500 ; 1 000[500	12	0,024
[1 000 ; 1 500[500	3	0,006
[1 500 ; 2 000[500	2	0,004

Ce qui permet de construire le diagramme suivant :



b) On calcule la **moyenne** :

$$m = \frac{5 \times 100 + 8 \times 350 + 12 \times 750 + 3 \times 1\,250 + 2 \times 1\,750}{30}$$

$$= 651,67$$

L'écart-type :

$$\begin{aligned}\sigma &= \sqrt{\frac{5 \times 100^2 + 8 \times 350^2 + 12 \times 750^2 + 3 \times 1\,250^2 + 2 \times 1\,750^2}{30}} - 651,67^2 \\ &= 441,68\end{aligned}$$

En comparaison avec les normes du groupe, il semble globalement que la production journalière moyenne est trop faible car la moyenne de 652 est inférieure à 1 000. De plus, il y a trop d'aléas dans la production, en effet l'écart-type, de l'ordre de 442, est supérieur à celui imposé par le groupe : 100.

c) Si on augmente la production journalière de 350 unités, la moyenne imposée par le groupe est atteinte.

Par contre, une augmentation de 350 unités par jour n'a aucun effet sur l'écart-type, par conséquent il se pose toujours le problème de l'aléa dans la production. Il est donc important de réguler l'augmentation, mais aussi la production de manière à niveler les écarts d'un jour à l'autre pour atteindre l'objectif d'écart-type faible.

CHAPITRE 2

Statistiques à deux variables

OBJECTIFS

- ▶ Établir l'existence d'un lien entre deux grandeurs
- ▶ Construire une droite de régression linéaire
- ▶ Connaître d'autres modes de régression
- ▶ Justifier l'utilisation d'une régression
- ▶ Utiliser l'outil de la régression pour faire des prévisions

PLAN

- 2.1 Covariance
- 2.2 Régression linéaire
- 2.3 Régressions non linéaires

En études de commerce, les grandeurs analysées sont liées entre elles : les couples quantités vendues-prix, capital-travail, rentabilité-risque, chiffre d'affaires-investissements, frais de publicité-ventes sont des exemples de grandeurs habituellement liées. Reste à savoir comment mesurer ce lien du point de vue de sa nature et de son intensité. Par-delà le constat de lien, de quelle manière l'évolution d'une grandeur influence-t-elle sur une autre ? Par exemple, quel effet une augmentation de prix aura-t-elle sur les quantités vendues ?

L'objectif de la modélisation est de réaliser des prévisions. La modélisation s'envisage sous deux approches :

- La première est **déterministe** : elle suppose l'existence théorique de lois et d'équations déterminées par l'économie ou la finance. Ces lois permettent de réaliser des prévisions. Cela suppose que le phénomène soit bien connu et qu'un modèle théorique efficient existe. Or, de nombreux phénomènes ne sont pas encore modélisés du fait de leur complexité.

- La seconde est **statistique** : en général, si on connaît mal le lien théorique entre les grandeurs, mais on dispose néanmoins d'observations. Ces observations servent à l'élaboration d'un modèle qui permet ensuite de faire des prévisions. Cette seconde approche est plus simple car les questions fondamentales de l'économie sont évitées, mais elle peut induire de fausses idées car les observations peuvent être le fruit du hasard et non d'une loi sous-jacente.

Dans ce chapitre et le suivant, la seconde approche sera détaillée, par conséquent l'ordre de grandeur des prévisions importera plus qu'un résultat précis.

2.1 COVARIANCE

Mode de représentation

Le mode de représentation d'une série à deux variables est le nuage de points. Le choix de la grandeur en abscisse et de la grandeur en ordonnée est *a priori* arbitraire sauf lorsqu'il est possible de préjuger d'un lien de cause à effet. Si la quantité vendue est la conséquence du prix de vente, on choisira de placer le prix de vente en abscisse et la quantité vendue en ordonnée.

La **covariance** est l'outil de mesure du lien entre deux grandeurs.

Mode de calcul



$$\text{Cov}(X, Y) = \frac{\sum_{i=1}^n x_i \times y_i - \sum_{i=1}^n x_i \times \sum_{i=1}^n y_i}{n^2}$$

avec x_i et y_i les valeurs des variables X et Y et n le nombre de points.



Le nombre de points n n'est pas le nombre de valeurs de X ajoutées au nombre de valeurs de Y mais uniquement le nombre de valeurs de X **ou** le nombre de valeurs de Y .

Une manière moins formelle de considérer cette formule est :

$$\text{Cov}(X, Y) = \text{Moyenne de } XY - \text{Moyenne de } X \times \text{Moyenne de } Y$$

Voyons sur des exemples le calcul et l'interprétation de la covariance :

Exemple.

Tableau 2-1

x_i	1	1	2	3	4
y_i	2	3	1	3	2

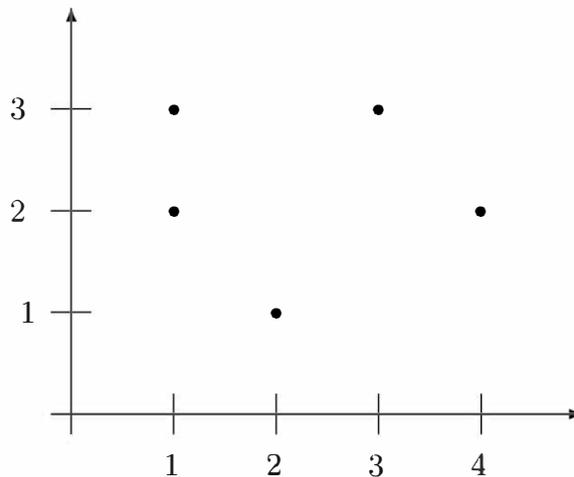


Figure 2-1

Calcul de covariance

$$\text{Cov}(x, y) = \frac{1 \times 2 + 1 \times 3 + 2 \times 1 + 3 \times 3 + 4 \times 2}{5} - \frac{1 + 1 + 2 + 3 + 4}{5} \times \frac{2 + 3 + 1 + 3 + 2}{5} = -0,04$$

On ne constate aucun lien graphique et une valeur quasi nulle de la covariance.

Exemple.

Tableau 2-2

x_i	1	1,5	2	3,5	4
y_i	1	2	2	2,5	3

Calcul de covariance

$$\text{Cov}(x, y) = \frac{1 \times 1 + 1,5 \times 2 + 2 \times 2 + 3,5 \times 2,5 + 4 \times 3}{5} - \frac{1 + 1,5 + 2 + 3,5 + 4}{5} \times \frac{1 + 2 + 2 + 2,5 + 3}{5} = 0,71$$

On constate que les variables X et Y évoluent ensemble (quand l'une augmente l'autre augmente aussi) et la valeur de la covariance est positive.

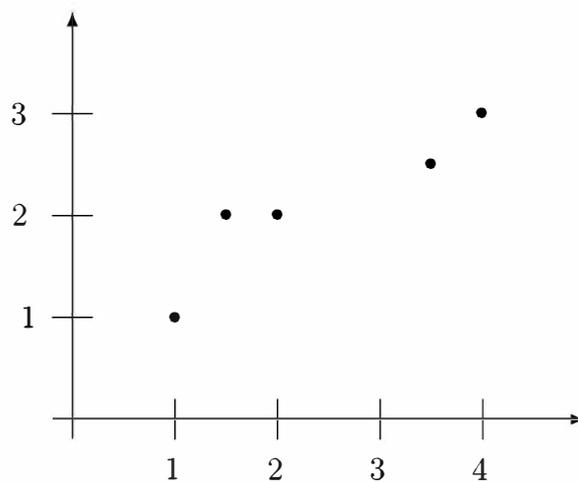


Figure 2-2

Exemple.

Tableau 2-3

x_i	1	2	2,5	3	4
y_i	3	2,5	2	1	0,5

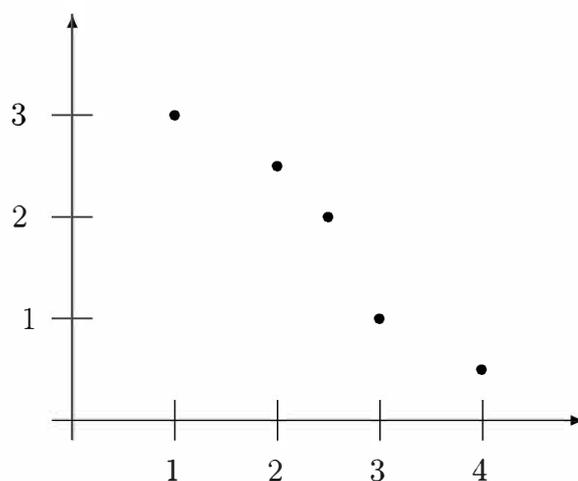


Figure 2-3

Calcul de covariance

$$\begin{aligned} \text{Cov}(x, y) &= \frac{1 \times 3 + 2 \times 2,5 + 2,5 \times 2 + 3 \times 1 + 4 \times 0,5}{5} - \frac{1 + 2 + 2,5 + 3 + 4}{5} \times \frac{3 + 2,5 + 2 + 1 + 0,5}{5} \\ &= -0,9 \end{aligned}$$

On constate que les variables X et Y évoluent en sens contraire (quand l'une augmente l'autre diminue) et la valeur de la covariance est négative.



Bilan

- Une covariance non nulle indique que les deux variables sont dépendantes.
- Si la covariance est positive, les variables évoluent dans le même sens.
- Si la covariance est négative, les variables évoluent en sens contraire.

Remarque : la covariance n'indique **que** le lien de dépendance, elle ne permet pas de dire si cette dépendance est linéaire, exponentielle ou autre.

Remarque : dans la pratique, on ne rencontre jamais ou presque de covariance nulle, mais des covariances quasi nulles. On considère qu'une covariance est quasi nulle si son ordre de grandeur est très faible devant l'ordre de grandeur des valeurs des variables. Ainsi une covariance de 100 sera quasi nulle si les valeurs de X et de Y sont de l'ordre du million.

2.2 RÉGRESSION LINÉAIRE

La manière la plus simple de modéliser une série est la régression linéaire. La régression linéaire consiste à assimiler un nuage de points à une droite. Cette droite est ensuite utile pour faire des prévisions.

Exemple. Voici une série à deux variables :

Tableau 2-4

x_i	1	2	3	4	5	6
y_i	1	2	4	2	4	4

Comment construire une droite assimilable à cette série ? Est-il légitime de le faire ? Quelle valeur de y_i correspondrait à $x_i = 7$?

a) Droite des points extrêmes

La droite des points extrêmes est la régression linéaire la plus simple.

Méthode : on assimile la série à la droite passant par le premier et le dernier point. On cherche son équation de manière à réaliser des prévisions.

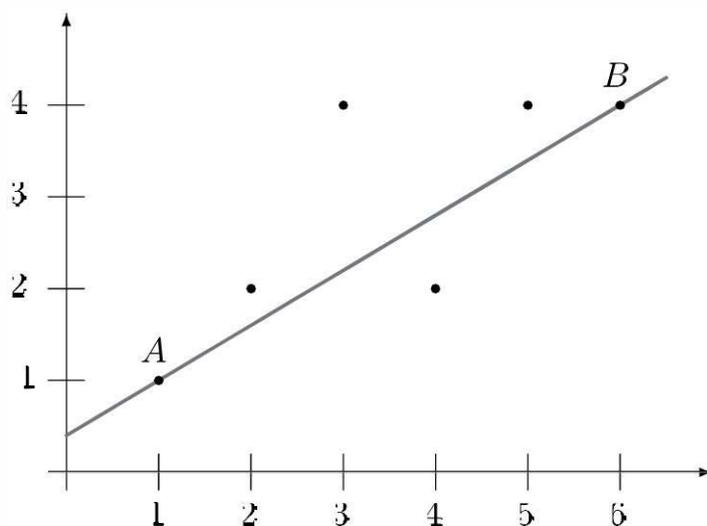


Figure 2-4

Le coefficient directeur de la droite (AB) est donné par la formule :

$$\frac{y_A - y_B}{x_A - x_B}$$

Application à l'exemple du Tableau 2-4 :

Les coordonnées des points extrêmes sont $A(1,1)$ et $B(6,4)$.

Le coefficient de la droite est donc $\frac{1 - 4}{1 - 6} = \frac{3}{5}$

L'équation réduite d'une droite affine est $y = ax + b$.

Ici $a = \frac{3}{5}$ et la droite passe par $A(1,1)$.

Ainsi $1 = \frac{3}{5} \times 1 + b$ donc $b = \frac{2}{5}$

L'équation de la droite est donc $y = \frac{3}{5}x + \frac{2}{5}$. On peut prévoir que si

$x = 7$, alors $y = \frac{3}{5} \times 7 + \frac{2}{5} = \frac{23}{5} = 4,6$

Remarque : cette méthode est la plus simple mais elle est imparfaite. Dans une observation, les points extrêmes peuvent être non représentatifs du reste de la série du fait d'un accident ou d'une erreur de mesure. La droite de régression ne sera alors pas représentative de la série.

b) Droite de Mayer

La droite de Mayer – encore appelée droite des points moyens – apporte une amélioration par rapport à la droite précédente car elle prend en compte l'ensemble des points de la série.

Méthode : on sépare la série en deux sous-séries ayant le même nombre de points. On détermine le point moyen de chacune des sous-séries et on détermine l'équation de la droite passant par ces deux points.

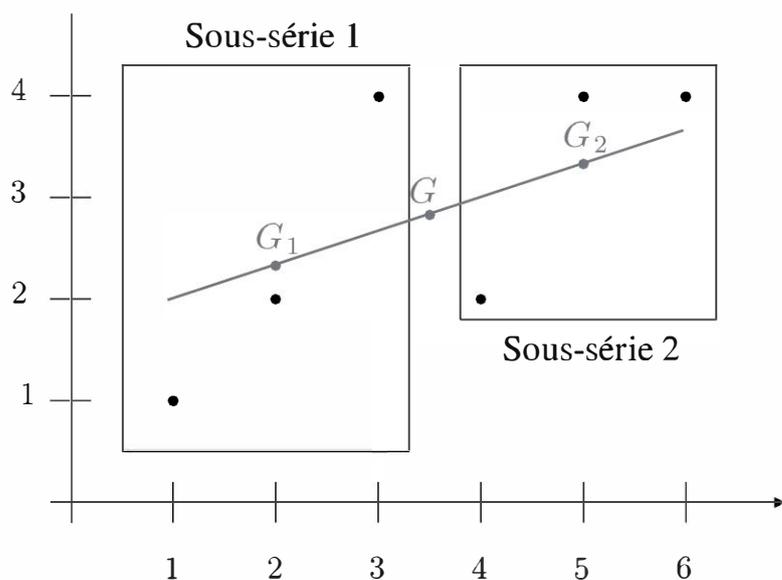


Figure 2-5

Application à l'exemple du Tableau 2-4

Notons G_1 et G_2 les points moyens des deux sous-séries. Les coordonnées de G_1 sont $\left(\frac{1+2+3}{3}; \frac{1+2+4}{3}\right) = \left(2; \frac{7}{3}\right)$

Les coordonnées de G_2 sont $\left(\frac{4+5+6}{3}; \frac{2+4+4}{3}\right) = \left(5; \frac{10}{3}\right)$

Le coefficient de la droite est donc $\frac{\frac{10}{3} - \frac{7}{3}}{5 - 2} = \frac{1}{3}$

L'équation réduite d'une droite affine est $y = ax + b$

Ici $a = \frac{1}{3}$ et la droite passe par $G_1 \left(2; \frac{7}{3}\right)$

Ainsi $\frac{7}{3} = \frac{1}{3} \times 2 + b$, donc $b = \frac{5}{3}$

L'équation de la droite de Mayer est $y = \frac{1}{3}x + \frac{5}{3}$. On peut prévoir ainsi que si $x = 7$, alors $y = \frac{1}{3} \times 7 + \frac{5}{3} = 4$.

Remarques

- La droite de Mayer passe par le point moyen $G(3,5; 2,83)$ de l'ensemble de la série.
- La série contient six points donc chacune des sous-séries a trois points. Le problème se pose si le nombre de points de la série est impair. Dans ce cas, il existe un point au centre de la série qui pourra aussi bien être dans la première sous-série que dans la seconde. On choisit de couper ce point en deux. Dans une série de 3 points, chacune des deux sous séries va contenir 1,5 points. Pour déterminer les coordonnées des points moyens, on devra pondérer les coordonnées du point au centre par 0,5.

c) Droite des moindres carrés

La droite des moindres carrés est la **droite optimale** ; elle est au plus proche des points de la série.

Méthode : l'équation de cette droite est $y = ax + b$ avec $a = \frac{\text{Cov}(X,Y)}{\text{Var}(X)}$ et $b = y_G - ax_G$, en notant $\text{Cov}(X,Y)$ la covariance de X et Y , $\text{Var}(X)$ la variance de X (ou l'écart-type de X au carré) et $(x_G; y_G)$ les coordonnées du point moyen de la série.

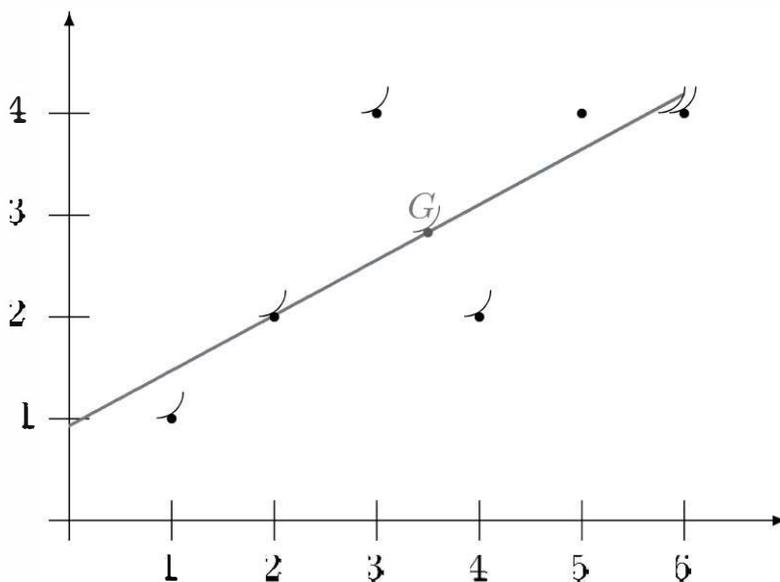


Figure 2-6

Application à l'exemple du Tableau 2-4

$$\begin{aligned} \text{On a } \text{Cov}(X, Y) &= \frac{1 \times 1 + 2 \times 2 + 3 \times 4 + 4 \times 2 + 5 \times 4 + 6 \times 4}{6} \\ &\quad - \frac{1 + 2 + 3 + 4 + 5 + 6}{6} \times \frac{1 + 2 + 4 + 2 + 4 + 4}{6} \\ &= \frac{19}{12} \end{aligned}$$

$$\begin{aligned} \text{Var}(X) &= \frac{1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2}{6} - \left(\frac{1 + 2 + 3 + 4 + 5 + 6}{6} \right)^2 \\ &= \frac{35}{12} \end{aligned}$$

$$\text{On a } a = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} = \frac{\frac{19}{12}}{\frac{35}{12}} = \frac{19}{35}$$

$$\text{On a } G \left(\frac{7}{2}; \frac{17}{6} \right) \text{ d'où } b = \frac{17}{6} - \frac{19}{35} \times \frac{7}{2} = \frac{14}{15}$$

L'équation de la droite des moindres carrés est donc $y = \frac{19}{35}x + \frac{14}{15}$

On peut prévoir ainsi que si $x = 7$ alors $y = \frac{19}{35} \times 7 + \frac{14}{15} \approx 4,73$

d) Coefficient de corrélation linéaire

Il reste à savoir s'il est légitime de faire une régression linéaire.

Pour savoir s'il est légitime d'effectuer une régression linéaire, on utilise le coefficient de **corrélation linéaire** : $r = \frac{\text{Cov}(X, Y)}{\sigma_X \times \sigma_Y}$ où σ_X et σ_Y sont les écart-types de X et de Y et $\text{Cov}(X, Y)$ la covariance.

Propriétés :

1. Si $r > 0$ (respectivement $r < 0$), la régression linéaire par la droite des moindres carrés conduit à une droite croissante (respectivement décroissante).
2. On a : $-1 \leq r \leq 1$
3. Si $r = 1$ ou $r = -1$, les points sont alignés. Ainsi, plus r est proche de 1 ou de -1 , plus la régression linéaire est appropriée.

2.3 RÉGRESSIONS NON LINÉAIRES

La régression linéaire est loin d'être le seul lien de dépendance entre deux variables, cependant les outils mis en place vont resservir pour les autres types de régressions. Dans chaque cas, il sera utile de revenir par une éventuelle transformation au cas de la régression linéaire. Ainsi la recherche d'une droite des moindres carrés, justifiée par une bonne valeur du coefficient de corrélation r , va permettre de trouver et de justifier de nouveaux types d'ajustements. Voici plusieurs exemples :

a) Ajustement exponentiel

L'ajustement exponentiel est un ajustement de la forme $y = ab^x$

Considérons l'exemple suivant :

Exemple.

Tableau 2-5

x_i	0,5	1	2	3	4
y_i	0,5	0,5	0,75	1,5	3,5

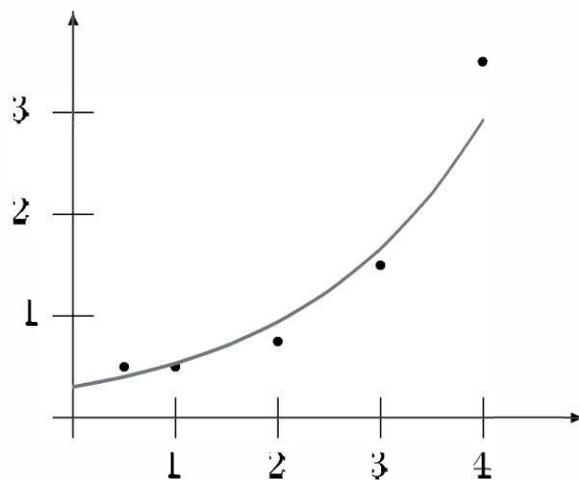


Figure 2-7

Le graphique ressemble davantage à une courbe de type exponentiel qu'à une droite. Comment justifier un tel choix ? Comment trouver les coefficients a et b qui définissent la régression exponentielle ? Dans la régression exponentielle $y = ab^x$ ainsi $\ln(y) = \ln(ab^x) \iff \ln(y) = \ln(a) + x \ln(b)$. Posons $y' = \ln(y)$, on a $y' = \ln(a) + x \ln(b)$. Une relation exponentielle

entre x et y équivaut donc à une relation affine entre x et y' . On peut ainsi justifier la régression exponentielle en calculant le coefficient r entre x et $\ln(y) = y'$:

Tableau 2-6

x_i	0,5	1	2	3	4
y'_i	-0,69	-0,69	-0,29	0,41	1,25

Par la formule précédente, on trouve $r = 0,97$. Ceci justifie le choix d'un ajustement exponentiel car r est proche de 1. En reprenant le calcul de la droite des moindres carrés, il est possible de retrouver a et b . On note que $\ln(b)$ est le coefficient directeur dans l'expression affine de y' en fonction de x et $\ln(a)$ est l'ordonnée à l'origine. Ainsi,

$$\ln(b) = \frac{\text{Cov}(X, Y')}{\text{Var}(X)} = 0,5665 \text{ et } \ln(a) = y'_G - 0,5665x_G = -1,1928.$$

On a donc $b = \exp(0,5665) = 1,7621$ et $a = \exp(-1,1928) = 0,3034$

D'où $y = 0,3034 \times 1,7621^x$

b) Ajustement logarithme

L'ajustement logarithme est de la forme $y = a \ln(x) + b$.

Considérons l'exemple suivant :

Exemple.

Tableau 2-7

x_i	0,5	1	2	3	4
y'_i	0,5	2	3	3,5	3,75

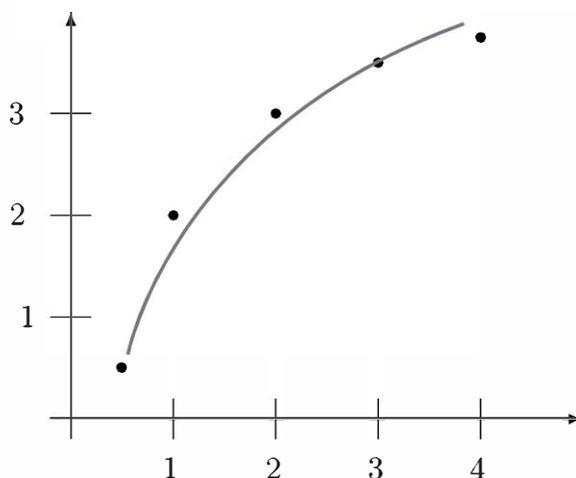


Figure 2-8

La répartition des points s'apparente à une courbe logarithmique. En posant $x' = \ln(x)$ on a : $y = ax' + b$. Ainsi, si x et y sont dans une dépendance logarithmique, x' et y sont dans une dépendance affine. Pour justifier le choix d'un ajustement logarithme, on calcule r entre x' et y . On détermine ensuite les coefficients a et b par la méthode des moindres carrés entre x' et y . On trouve par le calcul $r = 0,99$ d'où la légitimité de cet ajustement et $a = 1,555$ et $b = 1,777$.

Ainsi $y = 1,555 \ln(x) + 1,777$

c) Ajustement puissance

Le lien de dépendance entre x et y est : $y = ax^b$.

Ainsi $\ln(y) = \ln(a) + b \ln(x)$, la justification de cet ajustement se fera donc en calculant un coefficient de corrélation linéaire r entre $\ln(x)$ et $\ln(y)$. Les coefficients a et b se trouvent dans la recherche de la droite des moindres carrés pour $\ln(x)$ et $\ln(y)$.



POINTS CLEFS

- **La covariance** mesure la **dépendance** entre deux grandeurs. Elle se calcule ainsi :

$$\text{Cov}(X, Y) = \text{Moyenne des } XY - \text{Moyenne de } X \times \text{Moyenne de } Y$$

- **La corrélation linéaire** permet de déterminer si une régression linéaire est envisageable entre deux grandeurs. Elle se calcule ainsi :

$$r = \frac{\text{Cov}(X, Y)}{\sigma_X \times \sigma_Y}$$

Plus ce coefficient est proche de 1 ou de -1 meilleure sera la modélisation de la série par une droite.

- Il existe plusieurs façons de réaliser une régression linéaire ; la plus simple est la **droite des points extrêmes** qui consiste à assimiler la série à la droite passant par le premier et le dernier point. Une optimisation de ce modèle est la **droite de Mayer** que l'on obtient en divisant la série en deux sous-séries de même effectif pour déterminer la droite passant par les points moyens de ces deux sous-séries. La droite qui passe au plus proche de tous les points est la **droite des moindres carrés** :

$$\text{rés : } y = ax + b \text{ avec } a = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} \text{ et } b = y_G - a \times x_G$$

- Il existe d'autres régressions qui peuvent se justifier par des calculs de corrélations linéaires entre deux grandeurs. Voici quelques exemples :

Tableau 2-8

Type de régression	Formule de définition	Calcul de corrélation
Linéaire	$y = ax + b$	Entre X et Y
Puissance	$y = ax^b$	Entre $\ln(X)$ et $\ln(Y)$
Logarithme	$y = a \ln(x) + b$	Entre $\ln(X)$ et Y
Exponentielle	$y = ab^x$	Entre X et $\ln(Y)$

EXERCICES

2.1. Comptabilité analytique

En comptabilité analytique, on utilise le terme d'**unité d'œuvre** pour mesurer l'activité d'un centre. L'unité d'œuvre est une grandeur qui évolue en forte corrélation avec les coûts du centre. Dans l'exemple suivant, quelle unité d'œuvre doit-on choisir pour mesurer l'activité du centre ?

	Juillet	Août	Septembre	Octobre	Novembre	Décembre
Coût du centre	17 000	19 000	25 000	22 000	24 000	19 000
Heures de MOD	100	150	380	400	300	600
Heures-machines	600	1 800	3 600	3 000	3 000	1 200

2.2. CAC40

Le tableau suivant indique les rentabilités journalières en pourcentage du CAC40 et de l'action TF1 sur une période de 5 jours du 22/02/2010 au 26/02/2010.

CAC40	- 0,34	- 1,32	+ 0,23	- 2,02	+ 1,87
Action TF1	- 0,72	- 1,20	- 0,04	- 1,12	+ 2,93

a) Calculer la covariance de ces deux variables et le coefficient de corrélation linéaire. Le CAC40 est-il un indice pertinent pour l'étude du titre TF1 ?

b) En finance, on s'intéresse au coefficient $\beta = \frac{\text{Cov}(\text{Titre}, \text{CAC40})}{\text{Var}(\text{CAC40})}$.

Calculer ce coefficient pour le titre TF1. Qu'est-ce que ce résultat signifie pour un investisseur ?

2.3. Lien entre deux titres

Wendel est le principal actionnaire de Saint-Gobain. Dans cet exercice, on s'intéresse au lien entre les rentabilités journalières de ces deux titres sur la période du 22/02/2010 au 26/02/2010. Voici les résultats constatés en pourcentage :

Wendel	- 0,30	0,00	- 1,23	- 0,81	+ 6,84
Saint-Gobain	- 1,24	- 0,43	- 0,75	- 3,34	+ 7,88

a) Représenter graphiquement cette série.

b) Déterminer la droite de Mayer de la rentabilité de Wendel en fonction de celle de Saint-Gobain.

c) Si le cours Wendel progresse de 4 % le jour suivant, comment va évoluer le cours de Saint-Gobain selon ce modèle ?

2.4. Étude des ventes

On s'intéresse à la quantité vendue d'un produit dans un magasin en fonction de son prix. On a constaté les résultats suivants :

Prix de vente	5	10	15	25	35	45
Quantité vendue	140 000	35 000	15 500	5 500	3 000	1 500

a) Représenter graphiquement cette série.

b) Est-il plus pertinent de réaliser une régression linéaire ou une régression puissance ?

c) Réaliser la régression qui vous semble adaptée pour déterminer le prix de vente à fixer si on souhaite vendre 20 000 produits.

SOLUTIONS

Exercice 2.1

On s'intéresse à la grandeur qui a la plus forte corrélation avec les coûts du centre. Le calcul à réaliser est donc un coefficient de corrélation entre le coût du centre et l'une des deux grandeurs :

Pour l'heure de MOD : $Cov(Coût, MOD) =$

$$\frac{17\,000 \times 100 + 19\,000 \times 150 + 25\,000 \times 380 + 22\,000 \times 400 + 24\,000 \times 300 + 19\,000 \times 600}{6} - \frac{17\,000 + 19\,000 + 25\,000 + 22\,000 + 24\,000 + 19\,000}{6} \times \frac{100 + 150 + 380 + 400 + 300 + 600}{6} = 153\,333,33$$

$\sigma_{Coût} =$

$$\sqrt{\frac{17\,000^2 + 19\,000^2 + 25\,000^2 + 22\,000^2 + 24\,000^2 + 19\,000^2}{6} - \left(\frac{17\,000 + 19\,000 + 25\,000 + 22\,000 + 24\,000 + 19\,000}{6}\right)^2} = 2\,886,75$$

$$\sigma_{MOD} = \sqrt{\frac{100^2 + 150^2 + 380^2 + 400^2 + 300^2 + 600^2}{6} - \left(\frac{100 + 150 + 380 + 400 + 300 + 600}{6}\right)^2} = 166,38$$

$$\text{Ainsi on a : } r = \frac{Cov(Coût, MOD)}{\sigma_{Coût} \times \sigma_{MOD}} = \frac{153\,333,33}{2\,886,75 \times 166,38} = 0,3192$$

Pour l'heure-machine :

$Cov(Coût, Machine) =$

$$\frac{17\,000 \times 600 + 19\,000 \times 1\,800 + 25\,000 \times 3\,600 + 22\,000 \times 3\,000 + 24\,000 \times 3\,000 + 19\,000 \times 1\,200}{6} - \frac{17\,000 + 19\,000 + 25\,000 + 22\,000 + 24\,000 + 19\,000}{6} \times \frac{600 + 1\,800 + 3\,600 + 3\,000 + 3\,000 + 1\,200}{6}$$

$$= 3\,000\,000$$

$$\sigma_{Coût} = 2\,886,75$$

$\sigma_{Machine} =$

$$\sqrt{\frac{600^2 + 1\,800^2 + 3\,600^2 + 3\,000^2 + 3\,000^2 + 1\,200^2}{6} - \left(\frac{600 + 1\,800 + 3\,600 + 3\,000 + 3\,000 + 1\,200}{6}\right)^2}$$

$$= 1\,077,03$$

Ainsi on a :

$$r = \frac{Cov(Coût, Machine)}{\sigma_{Coût} \times \sigma_{Machine}} = \frac{3\,000\,000}{2\,886,75 \times 1\,077,03} = 0,9649$$

La corrélation entre le nombre d'heures-machines et les coûts du centre est supérieure à la corrélation entre le nombre d'heures de MOD et les coûts du centre. On retient donc **l'heure-machine** comme unité d'œuvre.

Exercice 2.2

a) On a $\text{Cov}(\text{CAC 40}, \text{TF1}) =$

$$\frac{-0,34 \times (-0,72) + (-1,32) \times (-1,20) + (0,23) \times (-0,04) + (-2,02) \times (-1,12) + 1,87 \times 2,93}{5} \\ = \frac{-0,34 - 1,32 + 0,23 - 2,02 + 1,87}{5} \times \frac{-0,72 - 1,20 - 0,04 - 1,12 + 2,93}{5} = 1,9027$$

La covariance est positive : les deux variables évoluent globalement dans le même sens. On observe sur ces 5 jours que lorsque le CAC40 baisse l'action TF1 baisse aussi et lorsque le CAC40 monte l'action TF1 monte aussi.

$$\sigma_{\text{CAC40}} = \sqrt{\frac{(-0,34)^2 + (-1,32)^2 + 0,23^2 + (-2,02)^2 + 1,87^2}{5} - \left(\frac{-0,34 - 1,32 + 0,23 - 2,02 + 1,87}{5}\right)^2} \\ = 1,3408$$

$$\sigma_{\text{TF1}} = \sqrt{\frac{(-0,72)^2 + (-1,20)^2 + (-0,04)^2 + (-1,12)^2 + 2,93^2}{5} - \left(\frac{-0,72 - 1,20 - 0,04 - 1,12 + 2,93}{5}\right)^2} = 1,5359$$

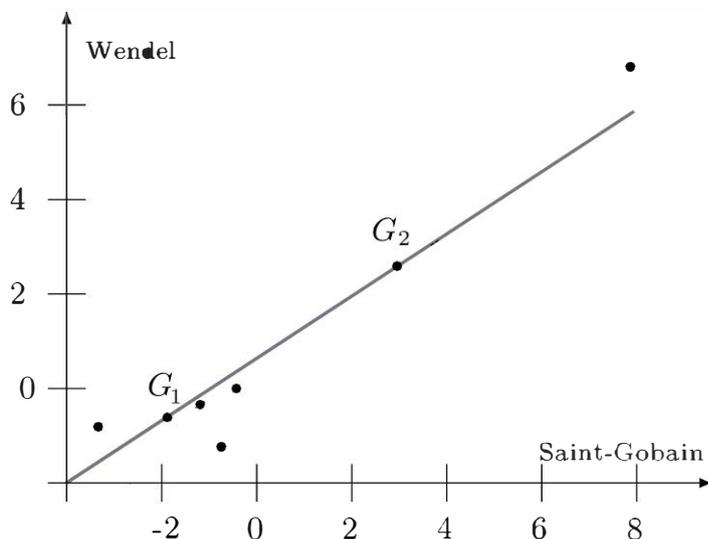
On a donc $r = \frac{1,9027}{1,3408 \times 1,5359} = 0,9239$. Ce résultat indique une forte corrélation entre les deux grandeurs. Le CAC40 est donc un indice pertinent pour l'étude de l'action TF1.

b) On a $\beta = \frac{1,9027}{1,3408^2} = 1,0584$. La formule définissant le coefficient β est

celle du coefficient directeur de la droite des moindres carrés de la rentabilité du titre TF1 en fonction de la rentabilité du CAC40. Ici $\beta = 1,06$, le titre évolue donc dans le sens du marché et amplifie le mouvement du marché. En effet, une hausse de 1 % de la rentabilité de l'indice entraîne une hausse de 1,06 % du titre. Le coefficient β fait apparaître un titre plutôt volatil et risqué sur la période. Notons qu'une observation de 5 jours est courte pour faire des prévisions ; cette conclusion serait à nuancer avec d'autres observations.

Exercice 2.3

a) Représentation graphique :



b) Dans un premier temps, classons les valeurs dans l'ordre croissant pour le titre Saint-Gobain :

Saint-Gobain	- 3,34	- 1,24	- 0,75	- 0,43	+ 7,88
Wendel	- 0,81	- 0,03	- 1,23	0	+ 6,84

La série contient 5 valeurs. Pour la méthode de la droite de Mayer, considérons deux sous-séries contenant chacune 2,5 points et recherchons les points moyens. Notons G_1 et G_2 les deux points moyens.

Les coordonnées de G_1 sont : $\left(\frac{-3,34-1,24-0,75 \times 0,5}{2,5} ; \frac{-0,81-0,3-1,23 \times 0,5}{2,5} \right)$

soit $(-1,982 ; -0,69)$. Notons que le coefficient 0,5 est associé au point du milieu.

Les coordonnées de G_2 sont $\left(\frac{-0,75 \times 0,5 - 0,43 + 7,88}{2,5} ; \frac{-1,23 \times 0,5 + 0 + 6,84}{2,5} \right)$,

soit $(2,83 ; 2,49)$.

Le coefficient directeur de la droite (G_1G_2) est $\frac{2,49 - (-0,69)}{2,83 - (-1,982)} = 0,661$.

L'ordonnée à l'origine de cette droite est donnée par les coordonnées de G_2 :

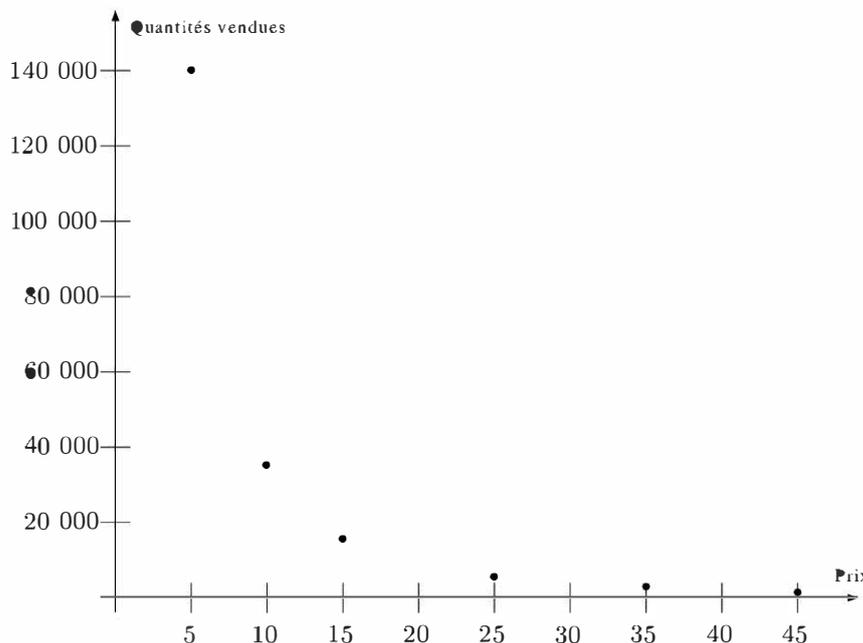
$$y_{G_2} - 0,661 \times x_{G_2} = 2,49 - 0,661 \times (2,83) = 0,619$$

L'équation de la droite de Mayer est donc : $y = 0,661x + 0,619$.

c) Si le cours de Wendel augmente de 4 % dans l'équation précédente, on a $y = 4$. Ainsi $4 = 0,661x + 0,619$ d'où $x = \frac{4 - 0,619}{0,661} = 5,11$. Si le titre Wendel augmente de 4 % selon cette modélisation le cours de Saint-Gobain va augmenter de 5,11 %.

Exercice 2.4

a) *Représentation graphique*



b) Dans un premier temps, calculons le coefficient de **corrélacion linéaire** entre les deux grandeurs :

$$\text{On a } \text{Cov}(Q, P) = \frac{140000 \times 5 + 35000 \times 10 + 15500 \times 15 + 5500 \times 25 + 3000 \times 35 + 1500 \times 45}{6} \\ - \frac{140000 + 35000 + 15500 + 5500 + 3000 + 1500}{6} \times \frac{5 + 10 + 15 + 25 + 35 + 45}{6} = -486458,33$$

La covariance est négative : les deux grandeurs évoluent en sens inverse.

$$\sigma_Q =$$

$$\sqrt{\frac{140000^2 + 35000^2 + 15500^2 + 5500^2 + 3000^2 + 1500^2}{6} - \left(\frac{140000 + 35000 + 15500 + 5500 + 3000 + 1500}{6}\right)^2} \\ = 49001,2$$

$$\sigma_P = \sqrt{\frac{5^2 + 10^2 + 15^2 + 25^2 + 35^2 + 45^2}{6} - \left(\frac{5 + 10 + 15 + 25 + 35 + 45}{6}\right)^2} = 14,07$$

On a donc $r = \frac{-486458,33}{49001,2 \times 14,07} = -0,7056$. La corrélation linéaire est donc moyenne.

Voyons maintenant l'opportunité d'une régression puissance. Pour cela, on doit déterminer le coefficient de corrélation linéaire entre les logarithmes des grandeurs :

ln(P)	1,60944	2,30259	2,70805	3,21888	3,55535	3,80666
ln(Q)	11,8494	10,4631	9,64860	8,61250	8,00637	7,31322

$$\text{Cov}(\ln(Q), \ln(P)) =$$

$$\frac{11,8494 \times 1,60944 + 10,4631 \times 2,30259 + 9,64860 \times 2,70805 + 8,61250 \times 3,21888 + 8,00637 \times 3,55535 + 7,31322 \times 3,80666}{6} \\ - \frac{11,8494 + 10,4631 + 9,64860 + 8,61250 + 8,00637 + 7,31322}{6} \times \frac{1,60944 + 2,30259 + 2,70805 + 3,21888 + 3,55535 + 3,80666}{6} \\ = -1,1529$$

$$\sigma_{\ln(Q)}$$

$$= \sqrt{\frac{11,8494^2 + 10,4631^2 + 9,64860^2 + 8,61250^2 + 8,00637^2 + 7,31322^2}{6} - \left(\frac{11,8494 + 10,4631 + 9,64860 + 8,61250 + 8,00637 + 7,31322}{6}\right)^2} \\ = 1,5311$$

$$\sigma_{\ln(P)}$$

$$= \sqrt{\frac{1,60944^2 + 2,30259^2 + 2,70805^2 + 3,21888^2 + 3,55535^2 + 3,80666^2}{6} - \left(\frac{1,60944 + 2,30259 + 2,70805 + 3,21888 + 3,55535 + 3,80666}{6}\right)^2} \\ = 0,7534$$

$$\text{On a donc } r = \frac{-1,1529}{1,5311 \times 0,7534} = -0,9994$$

La corrélation est donc très forte. La régression puissance semble donc significativement meilleure que la régression linéaire pour cette série.

c) Déterminons la droite des moindres carrés entre $\ln(Q)$ et $\ln(P)$. On a $\text{Cov}(\ln(Q), \ln(P)) = -1,1529$ et $\text{Var}(\ln(P)) = 0,7534^2 = 0,5676$.

Le coefficient directeur de la droite est donc : $\frac{-1,1529}{0,5676} = -2,0312$.

On trouve l'ordonnée à l'origine à l'aide du point moyen $G(2,8668 ; 9,3155)$. On a $y_G - (-2,0312) \times x_G = 15,1386$ comme ordonnée à l'origine. Ainsi, l'équation de la droite des moindres carrés est : $\ln(Q) = -2,0312 \times \ln(P) + 15,1386$

On a donc $Q = \exp(-2,0312 \times \ln(P) + 15,1386)$

soit $Q = \exp(-2,0312 \times \ln(P)) \times \exp(15,1386)$

d'où $Q = \exp(\ln(P^{-2,0312})) \times 3\,755\,004,34$

Ainsi $Q = 3\,755\,004,34 \times P^{-2,0312}$

Si on souhaite vendre 20 000 produits, on a :

$$20\,000 = 3\,755\,004,34 \times P^{-2,0312}$$

$$\text{soit } P^{-2,0312} = \frac{20\,000}{3\,755\,004,34} = 0,005326$$

Ainsi $P = 0,005\,326^{\frac{1}{-2,0312}} = 13,16$. Le prix de vente à fixer est donc de 13,16 euros si on souhaite vendre 20 000 produits.

CHAPITRE 3

Séries chronologiques

OBJECTIFS

- ▶ Déterminer la courbe de tendance centrale
- ▶ Calculer les moyennes mobiles
- ▶ Tracer les droites de support et de résistance pour déterminer la nature du modèle
- ▶ Utiliser la méthode des coefficients saisonniers pour faire des prévisions

PLAN

- 3.1 Moyennes mobiles
- 3.2 Résistance et support
- 3.3 Coefficients saisonniers
- 3.4 Phénomène de retracement

Les séries chronologiques sont des séries où une grandeur évolue en fonction du **temps**. Les variations observées ne correspondent pas toujours aux exemples étudiés au chapitre précédent :

- le sens de variation n'est pas constant ;
- des phénomènes saisonniers ou ponctuels apparaissent.

Considérons l'exemple suivant :

Exemple.

Tableau 3-1 Cours de clôture d'une action

Jour	0	1	2	3	4	5	6
Cours	2,5	2,6	2,5	2,8	2,7	3	2,9

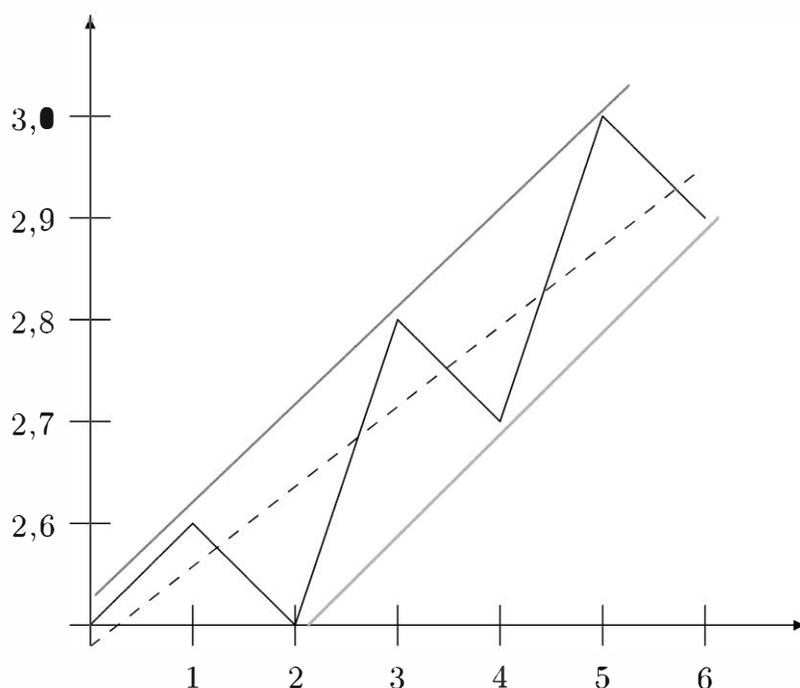


Figure 3-1

Modélisons cette série par une régression linéaire (droite des moindres carrés, en pointillés). Cette droite indique une évolution globale croissante. Cette droite sera appelée **tendance** ou *trend* pour une série chronologique. On constate que des variations positives ou négatives existent autour de la tendance. Il s'agit d'un phénomène dit **saisonnier**. La courbe des cours semble rebondir sur une droite imaginaire, cette droite est appelée **support** (droite en rose) et semble limitée en hauteur par une droite imaginaire appelée **résistance** (droite en rouge). La connaissance de la tendance, du support, de la résistance et du phénomène saisonnier va permettre de comprendre une évolution et de faire des prévisions.

Comment déterminer la tendance ? La droite des moindres carrés est une possibilité d'après l'exemple du *Tableau 3-1*. Les ajustements du chapitre précédents (puissance, exponentiel, logarithme) sont aussi envisageable dans certains cas. Comment justifier ces choix ? Les variations saisonnières diminuent la valeur du coefficient de corrélation linéaire. Son calcul est donc peu pertinent. On se limitera alors à une observation graphique pour choisir le bon ajustement.

Il arrive qu'aucun des modèles précédents ne soit pertinent pour déterminer la tendance. Cela arrive lorsque le mouvement global de la série n'est pas stable. Voyons avec les **moyennes mobiles** comment obtenir simplement une courbe de tendance sans compléxifier les modèles précédents.

3.1 MOYENNES MOBILES

Les moyennes mobiles servent à déterminer une tendance centrale dans une série chronologique. Les moyennes mobiles se calculent en effectuant la moyenne des valeurs autour du point considéré. Le nombre de valeurs prises en considération pour réaliser cette moyenne est la période.

Dans l'exemple suivant, on effectue le calcul des moyennes mobiles pour y_i avec différentes périodes.

Exemple.

Tableau 3-2 Moyennes mobiles

x_i	0	1	2	3	4	5	6
y_i	1	0	3	1	3	2	5
MM2		1	1,75	2	2,25	3	3
MM3		1,33	1,33	2,33	2	3,33	2,33
MM5			1,6	1,8	2,8	2,2	2,8
x_i	7	8	9	10	11	12	13
y_i	0	4	5	2	3	0	2
MM2	2,25	3,25	4	3	2	1,25	
MM3	3	3	3,67	3,33	1,67	1,67	
MM5	3,2	3,2	2,8	2,8	2,4		

Mode de calcul

- Pour la période 3 : la première valeur ne peut pas être calculée car on ne dispose pas d'un chiffre à sa gauche. On calcule donc la suivante. Les trois valeurs sont 1 ; 0 et 3, la moyenne mobile associée

est donc $\frac{1 + 0 + 3}{3} = 1,33$. Pour le calcul suivant, on coulisse la

sélection des trois valeurs d'un cran, les valeurs à considérer sont donc : 0 ; 3 ; 1 et la moyenne mobile est 1,33 à nouveau. On continue ainsi de suite.

- Pour la période 5 : les deux premières valeurs ne sont pas calculables car il manque deux valeurs à gauche. La troisième valeur se calcule en effectuant la moyenne de 1 ; 0 ; 3 ; 1 ; 3. On trouve

$\frac{1 + 0 + 3 + 1 + 3}{5} = 1,6$. Pour la suivante on considère les cinq

valeurs qui suivent en décalant d'un cran : 0 ; 3 ; 1 ; 3 ; 2. D'où une moyenne de 1,8. On continue ainsi de suite.

- Pour la période 2 : le problème va se poser plus généralement lorsque la période est paire. Pour centrer autour de la valeur considérée, on coupe les valeurs limites en deux. Par exemple, pour le premier calcul les valeurs considérées sont 1 ; 0 et 3. La valeur centrale est 0, on la compte donc une fois, par contre les deux valeurs limites 1 et 3 seront comptés une demi fois de manière à avoir un total de 2 points. Le calcul est donc :
$$\frac{0,5 \times 1 + 0 + 0,5 \times 3}{2} = 1$$

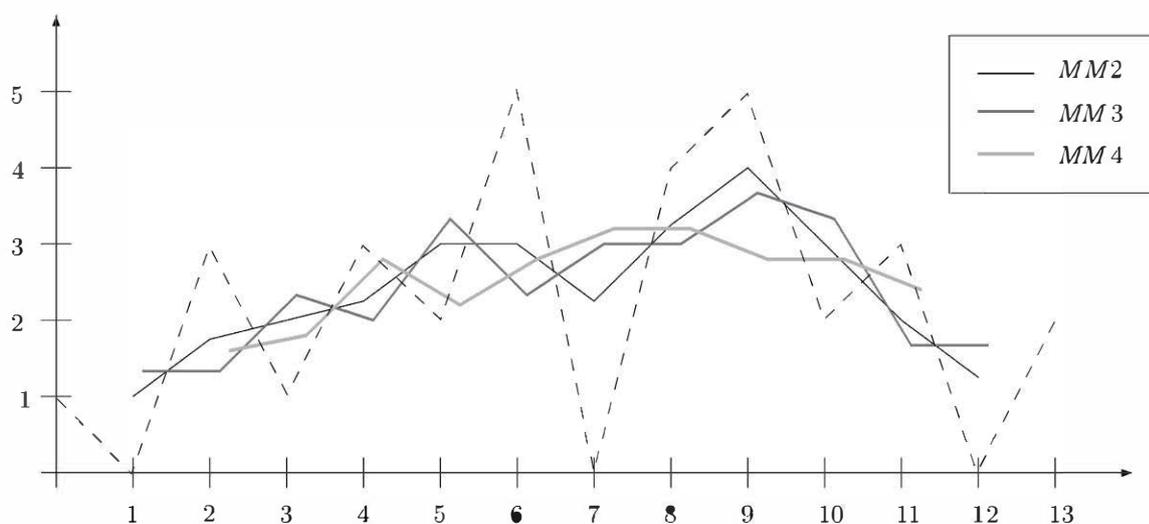


Figure 3-2 Moyennes mobiles

On constate que plus la période est grande, moins les moyennes mobiles varient. En effet, les moyennes mobiles de période 2 évoluent entre 1 et 4, soit une amplitude de 3 ; les moyennes mobiles de période 3 évoluent entre 1,33 et 3,67, soit une amplitude de 2,34, et les moyennes mobiles de période 5 évoluent entre 1,6 et 3,2, soit une amplitude de 1,6.

Les moyennes mobiles procèdent à un **lissage** des valeurs qui augmente avec la période. Le choix de la période dépend de l'importance du lissage souhaité :

- Période trop petite \implies Prise en compte trop importante des évolutions ponctuelles \implies Moyennes mobiles qui varient fortement.
- Période trop grande \implies Non prise en compte de l'évolution globale \implies Moyennes mobiles qui ne varient pas.

Remarque : les moyennes mobiles permettent de lisser une courbe et donc d'en simplifier la lecture par contre, à l'inverse des régressions du chapitre précédent, elles ne permettent pas de faire des prévisions.

3.2 RÉSISTANCE ET SUPPORT

La résistance et le support sont les droites qui encadrent la série.

a) Mode de construction des droites

On peut construire simplement la droite de résistance (ou de support) lorsque au moins trois maximums (ou minimums) sont alignés. En leur absence, on peut envisager la méthode d'**Andrews' Pitchfork**.



Méthode d'Andrews' Pitchfork

Cette méthode s'utilise lorsque le mouvement est globalement haussier avec un retracement (une petite baisse) ou globalement baissier avec un retracement (une petite hausse). Considérons le premier cas :

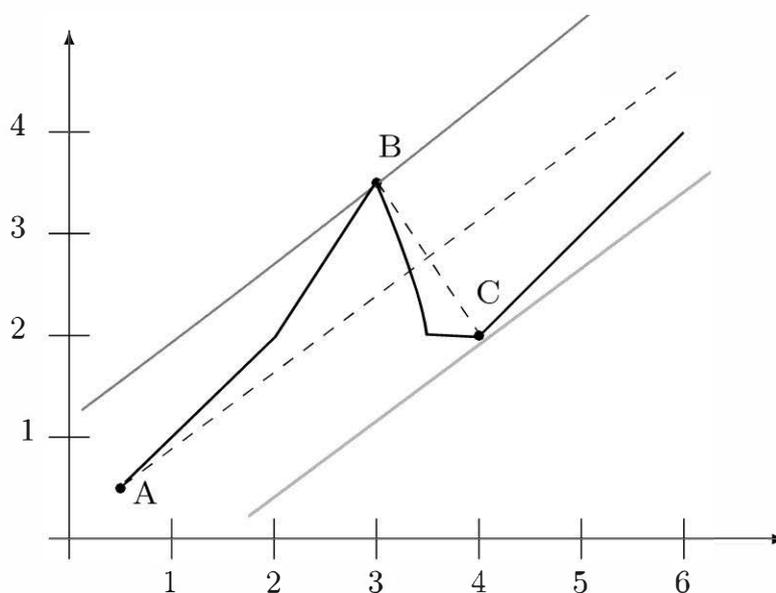


Figure 3-4

On place un point A à la base de la courbe, un point B avant le retracement et un point C à la fin du retracement. On trace la droite passant par A et le milieu de [BC]. La parallèle à cette droite passant par B est la résistance (en rouge) et la parallèle à cette droite passant par C est le support (en rose).

b) Modèle additif et modèle géométrique

On peut constater différents positionnement de la droite support et de la droite de résistance.

Le modèle est dit **additif** lorsque le support et la résistance sont parallèles. Dans ce cas, l'amplitude des oscillations demeure constante.

Lorsque la résistance et le support se coupent, le modèle est dit **multipliatif**. Les deux droites se coupent en amont ou en aval de la courbe. En aval, l'amplitude des oscillations diminue, en amont l'amplitude des oscillations augmente.

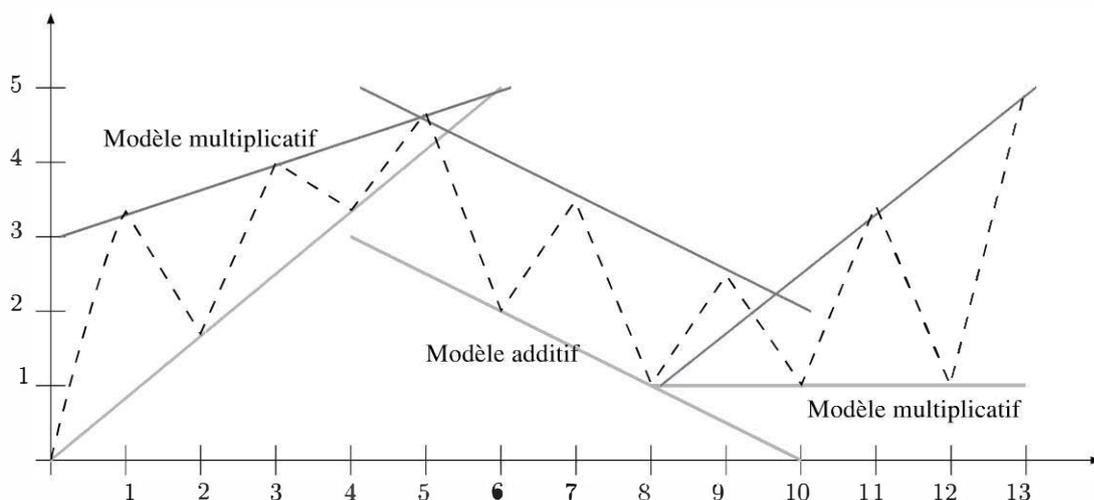


Figure 3-3 Position relative de la résistance et du support

Remarque : une série ne reste jamais perpétuellement dans le même modèle, un percement du support ou de la résistance peut indiquer un changement de modèle dans l'étude d'un cours de bourse.

3.3 COEFFICIENTS SAISONNIERS

a) Pour un modèle additif

Le modèle additif se caractérise par une relation additive entre la tendance centrale et le phénomène saisonnier. On a $y_i = t_i + s_i$ où t_i est la tendance que l'on détermine à l'aide d'une des méthodes de régression et s_i est un coefficient saisonnier.

Considérons l'exemple suivant :

Exemple.

Tableau 3-3

x_i	0	1	2	3	4	5
y_i	3	4	2,5	3,5	1,5	2,5

La droite des moindres carrés de cette série est :

$$y = -0,2571 \times x + 3,4762. \text{ Cette droite est la tendance } t_i$$

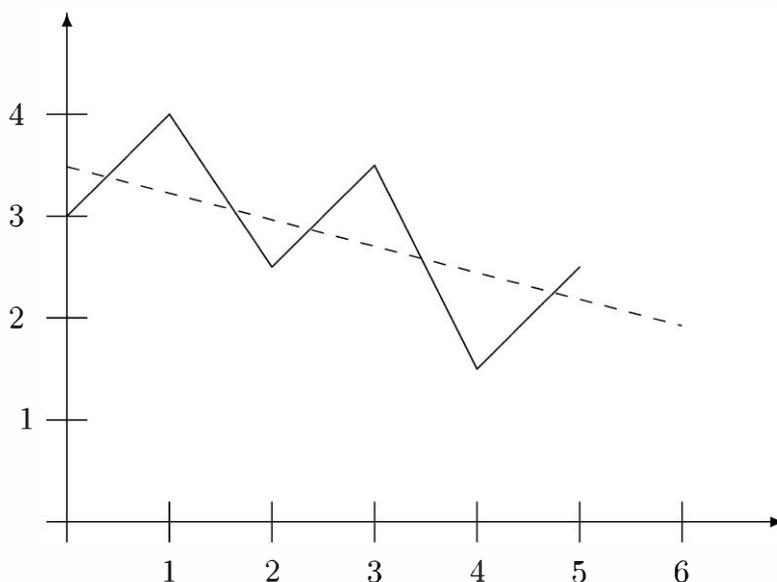


Figure 3-5

Tableau 3-4 Écart entre tendance et valeurs

x_i	y_i	t_i	$y_i - t_i$
0	3	3,4762	- 0,4762
1	4	3,2191	0,7809
2	2,5	2,962	- 0,462
3	3,5	2,7049	0,7951
4	1,5	2,4478	- 0,9478
5	2,5	2,1907	0,3093

Le *Tableau 3-4* renseigne sur l'écart entre la tendance et les valeurs réelles de la série. On constate que cette série a deux saisons caractérisées alternativement par un écart à la tendance positif et un écart à la tendance négatif. On peut supposer que cette habitude de la série va se poursuivre. Pour réaliser une prévision, il serait bon de déterminer un écart moyen entre la tendance et les valeurs selon la saison.

– Pour la saison positive l'écart moyen est de :

$$\frac{0,7809 + 0,7951 + 0,3093}{3} = 0,6284$$

– Pour la saison négative l'écart moyen est de :

$$\frac{-0,4762 - 0,462 - 0,9478}{3} = -0,6287$$

Ces deux valeurs seront admises comme coefficients saisonniers : s_i

Remarque : la somme des écarts à la tendance est quasi nulle donc la somme des coefficients saisonniers est quasi nulle aussi.

Prévisions :

- Pour $x_i = 6$: on a $t_i = -0,2571 \times 6 + 3,4762 = 1,9336$ d'après la droite des moindres carrés. La période 5 était positive par conséquent la suivante sera négative, ainsi $s_i = -0,6287$. La prévision est donc : $y_i = 1,9336 - 0,6287 = 1,3049$.
- Pour $x_i = 7$: on a $t_i = -0,2571 \times 7 + 3,4762 = 1,6765$ d'après la droite des moindres carrés. La période 6 était négative par conséquent la suivante sera positive, ainsi $s_i = 0,6284$. La prévision est donc : $y_i = 1,6765 + 0,6284 = 2,3049$.

b) Pour un modèle multiplicatif

Le modèle multiplicatif se caractérise par une **relation multiplicative entre la tendance et le phénomène saisonnier**. On a $y_i = t_i \times s_i$ où t_i est la tendance que l'on détermine à l'aide d'une des méthode de régression et s_i est un coefficient saisonnier.

Considérons l'exemple suivant :

Exemple.

Tableau 3-5

x_i	1	2	3	4	5	6
y_i	0,5	2	0,5	2,5	1	4

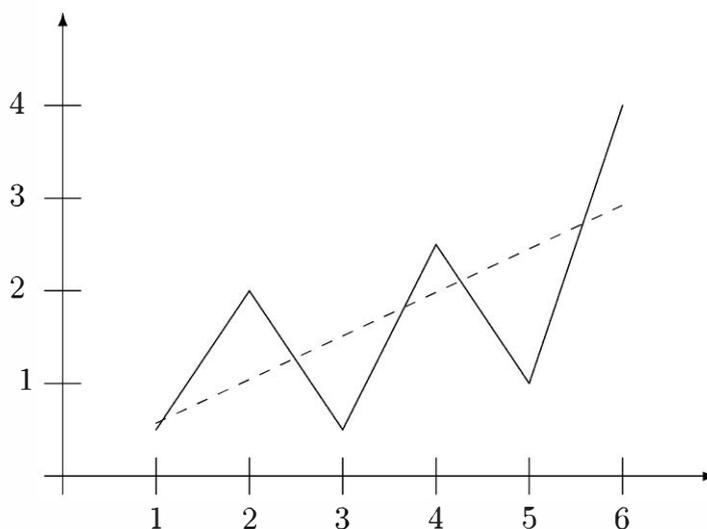


Figure 3-6

La droite des moindres carrés de cette série est : $y = 0,4714 \times x + 0,1$
 Cette droite est la tendance t_i

Tableau 3-6 Rapport entre valeurs et tendance

x_i	y_i	t_i	$\frac{y_i}{t_i}$
1	0,5	0,5714	0,8750
2	2	1,0428	1,9179
3	0,5	1,5142	0,3302
4	2,5	1,9856	1,2591
5	1	2,457	0,407
6	4	2,9284	1,3659

De la même manière que dans la série précédente, on constate alternativement une saison positive suivie d'une saison négative.

– Pour la saison positive, le coefficient saisonnier moyen est :

$$\frac{1,9179 + 1,2591 + 1,3659}{3} = 1,5143$$

– Pour la saison négative, le coefficient saisonnier moyen est :

$$\frac{0,8750 + 0,3302 + 0,407}{3} = 0,5374$$

Remarque : la somme des coefficients moyens est proche de 2. De manière générale, la somme des coefficients saisonniers sera proche du nombre de saisons.

Prévisions :

– Pour $x_i = 7$: on a $t_i = 0,4714 \times 7 + 0,1 = 3,3998$ d'après la droite des moindres carrés. La période 6 était positive par conséquent la suivante sera négative, ainsi $s_i = 0,5374$. La prévision est donc :

$$y_i = 3,3998 \times 0,5374 = 1,8271$$

– Pour $x_i = 8$: on a $t_i = 0,4714 \times 8 + 0,1 = 3,8712$ d'après la droite des moindres carrés. La période 7 était négative par conséquent la suivante sera positive, ainsi $s_i = 1,5143$. La prévision est donc :

$$y_i = 3,8712 \times 1,5143 = 5,8622$$

3.4 PHÉNOMÈNE DE RETRACEMENT

Certaines séries chronologiques évoluent pendant une période donnée dans un même sens.

Considérons l'exemple suivant :

Exemple.

Tableau 3-7

x_i	0	1	2	3	4	5	6
y_i	3,5	3,7	3,7	3,9	4	4,4	4,5

Le cours de l'action ne fait qu'augmenter, le mouvement est haussier, pourtant il est important d'envisager l'éventualité d'un retracement (une légère baisse).

Les **ratios de Fibonacci** : 38 %, 50 % et 62 % permettent d'envisager différents scénarios de retracement.

La formule pour trouver la valeur après retracement est :

- Dernier cours – Amplitude de la hausse \times Ratio dans le cas d'un mouvement haussier
- Dernier cours + Amplitude de la baisse \times Ratio dans le cas d'un mouvement baissier.

Ici le dernier cours est 4,5, l'amplitude est $4,5 - 3,5 = 1$. Les différents retracements envisageables sont :

- Pour un ratio de 38 % : $4,5 - 0,38 \times 1 = 4,12$
- Pour un ratio de 50 % : $4,5 - 0,5 \times 1 = 4$
- Pour un ratio de 62 % : $4,5 - 0,62 \times 1 = 3,88$



POINTS CLEFS

- Une série chronologique se décompose en une **tendance**, une droite de **support** et une droite de **résistance**.
- La **tendance** se détermine à l'aide d'une régression ou par la méthode des **moyennes mobiles**.

- Les **moyennes mobiles** lissent les valeurs de la série. L'importance du lissage augmente avec la période.
- Les droites de **support** et de **résistance** passent au moins par trois points extrêmes de la courbe. Sinon, la méthode de **Pitchfork** permet de construire ces droites.
- Il existe deux modèles de séries chronologiques, le **modèle additif** pour lequel support et résistance sont parallèles et le **modèle multiplicatif** pour lequel ces deux droites sont sécantes.
- Pour le **modèle additif**, le calcul de l'écart moyen par période entre la tendance et la série permet de calculer des coefficients saisonniers utiles pour faire des prévisions.
- Pour le **modèle multiplicatif**, le calcul du rapport moyen par période entre la série et la tendance permet de déterminer des coefficients saisonniers utiles pour faire des prévisions.
- Lorsqu'un mouvement est globalement haussier ou baissier, on peut envisager un **retracement** (changement de sens de variation) à l'aide des **ratios de Fibonacci**.

EXERCICES

3.1 Chiffre d'affaires dans le tourisme

On s'intéresse au chiffre d'affaires d'un hôtel en fonction du trimestre sur plusieurs années. Voici les valeurs constatées en milliers d'euros :

	2006	2007	2008	2009
Trimestre 1	30	35	38	40
Trimestre 2	45	50	52	56
Trimestre 3	150	160	165	180
Trimestre 4	80	90	95	100

- a) Représenter graphiquement cette série.
- b) De quel modèle s'agit-il ?
- c) Calculer les moyennes mobiles de période 4 de cette série.

- d) Représenter graphiquement les moyennes mobiles.
- e) Calculer les coefficients saisonniers et interpréter les valeurs trouvées.

3.2 Cours de clôture d'une action

On s'intéresse au cours de clôture d'une action sur plusieurs semaines boursières :

	Lundi	Mardi	Mercredi	Jeudi	Vendredi
Semaine 1	30	32	33	35	32
Semaine 2	31	34	36	40	35
Semaine 3	32	36	39	50	42

- a) Représenter graphiquement cette série.
- b) De quel modèle s'agit-il ?
- c) Déterminer la droite des moindres carrés de cette série.
- d) Calculer les coefficients saisonniers.
- e) Réaliser une prévision pour les cours de la semaine 4.

3.3 Estimations

Voici le cours d'une action sur plusieurs jours :

Cours	10	9,5	8	7	9	8	7,5	6,5
Jours	1	2	3	4	5	6	7	8

- a) Représenter graphiquement cette série.
- b) Déterminer par la méthode de Pitchfork les équations des droites de résistance et de support.
- c) En déduire un intervalle probable dans lequel va se trouver l'action le jour 10.
- d) Avec un ratio de 50 %, quel retracement peut-on envisager ?

3.4 Achat ou vente de titre ?

En finance de marché, pour un titre donné, les moyennes mobiles de période n sont calculées par les moyennes des n dernières valeurs connues. Ceci donne un poids important aux données passées et permet de recalculer les moyennes mobiles chaque jour sans avoir besoin de connaître les cours des jours suivants (qui sont de toute évidence

inconnus). Les signaux associés aux calculs des moyennes mobiles sont les suivants :

- acheter en phase de hausse, quand le cours traverse la moyenne mobile de bas en haut ;
- vendre en phase de baisse, quand le cours traverse la moyenne mobile de haut en bas.

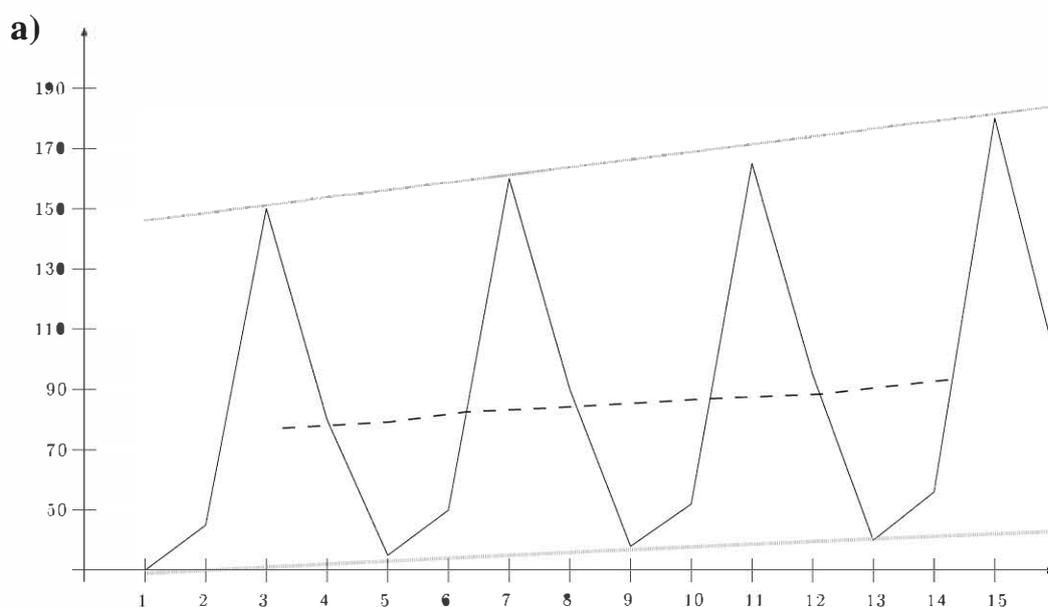
On a relevé les cours d'une action sur 20 jours consécutifs.

Jour	1	2	3	4	5	6	7	8	9	10
Cours	101	101,3	100,8	101,6	102	101,8	100,9	102,5	103,2	101,8
Jour	11	12	13	14	15	16	17	18	19	20
Cours	102,4	103,9	105	108	110	106,5	105,7	106	109,5	112

- a) Calculer les moyennes mobiles de période 4 et de période 8.
- b) Utiliser les moyennes mobiles et la règle précédente pour déterminer les dates et les opérations intéressantes à effectuer.

SOLUTIONS

Exercice 3.1



- b) Un tracé rapide des droites de résistance et de support permet de constater que le modèle est additif. En effet, sur la période considérée, les deux droites sont quasiment parallèles.

c) Les deux premières valeurs ne sont pas calculables, la troisième est la moyenne des quatre valeurs qui l'entourent. Les deux valeurs du bord sont comptées une demi fois. La moyenne mobile est donc :

$$\frac{0,5 \times 30 + 45 + 150 + 80 + 35 \times 0,5}{4} = 76,875.$$

Le calcul suivant est

$$\frac{0,5 \times 45 + 150 + 80 + 35 + 0,5 \times 50}{4} = 78,125.$$

Les résultats des calculs sont présentés dans le tableau suivant :

	2006	2007	2008	2009
Trimestre 1		80	85,625	90,875
Trimestre 2		82,5	86,875	93,375
Trimestre 3	76,875	84,125	87,75	
Trimestre 4	78,125	84,75	88,5	

Remarque : la période 4 est un bon choix puisque les variations saisonnières sont liées aux trimestres.

d) Tracé des moyennes mobiles sur le graphique de la page précédente. On constate une courbe (en pointillés) qui forme quasiment une droite représentant la tendance de la série.

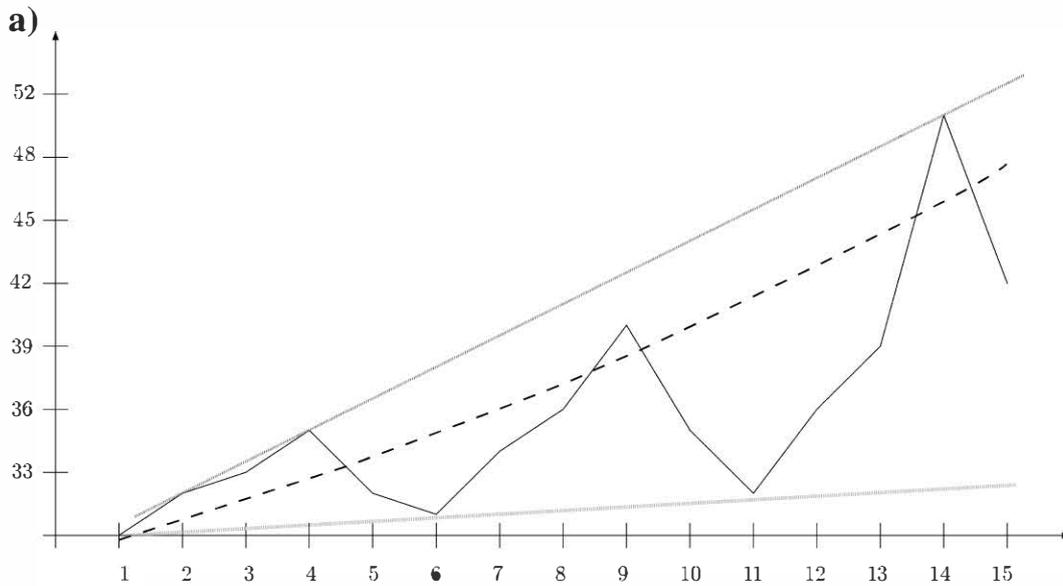
e) Calcul des coefficients saisonniers.

Le tableau suivant donne l'écart entre la série et la tendance, la dernière colonne est le coefficient trimestriel moyen.

	2006	2007	2008	2009	Coefficient trimestriel
Trimestre 1		-45	-47,625	-50,875	-47,833
Trimestre 2		-32,5	-34,875	-37,375	-34,817
Trimestre 3	73,125	75,875	77,25		75,417
Trimestre 4	1,875	5,25	6,5		4,542

Le premier trimestre est le pire de l'année, le troisième trimestre est le meilleur, le quatrième est plutôt neutre et le second est plutôt mauvais. On peut comprendre ces résultats dans le tourisme car le chiffre d'affaires réalisé est sensiblement plus important dans les périodes de vacances. On peut supposer que ces écarts vont se maintenir en 2010.

Exercice 3.2



b) Un tracé rapide des droites de résistance et de support permet de constater que le modèle est multiplicatif.

c) Pour déterminer la droite des moindres carrés de la série, on associe à chaque cours le jour correspondant. L'équation de la droite est : $y = 0,8571x + 28,9429$. On note que le coefficient de corrélation linéaire est de 0,73 ce qui est moyen, d'où la nécessité de prendre en compte les variations saisonnières.

d) Pour calculer les coefficients saisonniers, calculons dans un premier temps le rapport $\frac{y_i}{t_i}$ où t_i est donné par la droite de tendance ($y = 0,8571x + 28,9429$) et y_i est le cours de l'action :

x_i	y_i	t_i	$\frac{y_i}{t_i}$	x_i	y_i	t_i	$\frac{y_i}{t_i}$
1	30	29,8	1,0067	9	40	36,657	1,0912
2	32	30,657	1,0438	10	35	37,514	0,9330
3	33	31,514	1,0472	11	32	38,371	0,8340
4	35	32,371	1,0812	12	36	39,229	0,9177
5	32	33,229	0,9630	13	39	40,086	0,9729
6	31	34,086	0,9095	14	50	40,943	1,2212
7	34	34,943	0,9730	15	42	41,8	1,0048
8	36	35,8	1,0056				

On constate graphiquement que l'effet saisonnier est lié au jour de la semaine. Un motif semble se reproduire d'une semaine à l'autre. On choisit donc de calculer un coefficient par jour de la semaine.

– Pour le lundi, il s'agit des jours numérotés 1, 6 et 11. Le coefficient du lundi est donc :

$$\frac{1,0067 + 0,9095 + 0,8340}{3} = 0,9167$$

– Pour le mardi, il s'agit des jours numérotés 2, 7 et 12. Le coefficient du mardi est donc :

$$\frac{1,0438 + 0,9730 + 0,9177}{3} = 0,9782$$

– Pour le mercredi, il s'agit des jours numérotés 3, 8 et 13. Le coefficient du mercredi est donc :

$$\frac{1,0472 + 1,0056 + 0,9729}{3} = 1,0086$$

– Pour le jeudi, il s'agit des jours numérotés 4, 9 et 14. Le coefficient du jeudi est donc :

$$\frac{1,0812 + 1,0912 + 1,2212}{3} = 1,1312$$

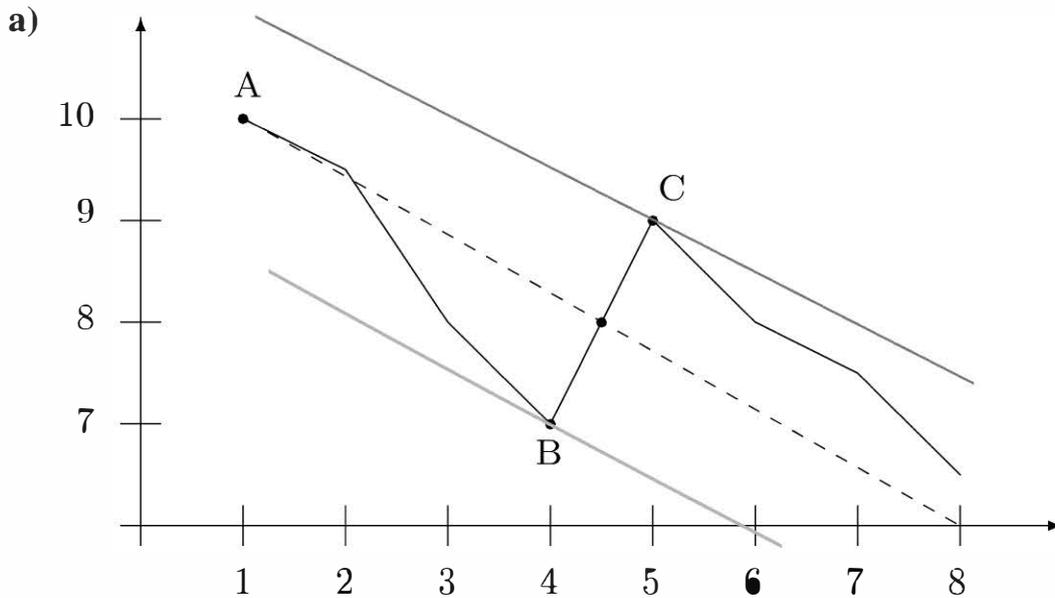
– Pour le vendredi, il s'agit des jours numérotés 5, 10 et 15. Le coefficient du vendredi est donc :

$$\frac{0,9630 + 0,9330 + 1,0048}{3} = 0,9669$$

e) Pour la semaine 4, on réalise une estimation de la tendance à l'aide de la droite des moindres carrés et grâce aux coefficients saisonniers, on corrige la tendance :

x_i	t_i	Coefficient	Estimation
16	42,657	0,9167	39,1037
17	43,514	0,9782	42,5654
18	44,371	1,0086	44,7526
19	45,229	1,1312	51,1630
20	46,086	0,9669	44,5606

Exercice 3.3



b) On place les points $A(1 ; 10)$, $B(4 ; 7)$ et $C(5 ; 9)$. Le milieu de $[BC]$ est de coordonnées $(4,5 ; 8)$. La droite qui passe par A et le milieu de $[BC]$ a pour coefficient directeur : $\frac{8 - 10}{4,5 - 1} = -\frac{4}{7} = -0,5714$.

La résistance est la parallèle à cette droite passant par C . Ainsi, le coefficient directeur de la résistance est aussi $-\frac{4}{7}$. L'ordonnée à l'origine est : $9 - 5 \times \left(-\frac{4}{7}\right) = \frac{83}{7} = 11,8571$. Ainsi, l'équation de la résistance est $y = -\frac{4}{7} \times x + \frac{83}{7}$

Le raisonnement est le même pour le support : le support passe par B , son ordonnée à l'origine est donc : $7 - 4 \times \left(-\frac{4}{7}\right) = \frac{65}{7} = 9,2857$

L'équation du support est donc $y = -\frac{4}{7} \times x + \frac{65}{7}$

c) Le jour 10, le cours de l'action va se situer *a priori* entre le support et la résistance.

Le jour 10, le support est de $-\frac{4}{7} \times 10 + \frac{65}{7} = \frac{25}{7} = 3,5714$, la résistance est de $-\frac{4}{7} \times 10 + \frac{83}{7} = \frac{43}{7} = 6,1429$. L'intervalle dans lequel se situera le cours de l'action est donc $[3,5714 ; 6,1429]$.

d) Le mouvement est globalement baissier, un retracement correspond ici à une remontée des cours. Depuis le point A, l'amplitude de la baisse est de $10 - 6,5 = 3,5$. Avec un ratio de 50 %, on peut envisager une remontée du cours à $6,5 + 3,5 \times 50 \% = 8,25$.

Exercice 3.4

a) Calcul des moyennes mobiles :

Jour	1	2	3	4	5	6	7	8
Cours	101	101,3	100,8	101,6	102	101,8	100,9	102,5
MM4				101,18	101,43	101,55	101,58	101,8
MM8								101,49

Jour	9	10	11	12	13	14	15	16
Cours	103,2	101,8	102,4	103,9	105	108	110	106,5
MM4	102,1	102,1	102,48	102,83	103,28	104,83	106,73	107,38
MM8	101,76	101,83	102,03	102,31	102,69	103,46	104,6	105,1

Jour	17	18	19	20
Cours	105,7	106	109,5	112
MM4	107,55	107,05	106,93	108,3
MM8	105,41	105,94	106,93	107,84

On note que dans cet exercice les moyennes mobiles ne sont pas calculées comme dans le cours. Elles sont calculées soit avec les quatre valeurs précédentes, soit avec les huit valeurs précédentes.

Par exemple, pour la première moyenne mobile de période 4, on trouve :

$$\frac{101 + 101,3 + 100,8 + 101,6}{4} = 101,18$$

b) Utilisation des moyennes mobiles pour donner un conseil.

Moyennes mobiles de période 4 :

- Vendre le jour 7 à 100,9 euros (les moyennes mobiles passent au dessus des cours).
- Acheter le jour 8 à 102,5 euros (les moyennes mobiles passent en dessous des cours).
- Vendre le jour 10 à 101,8 euros.

- Acheter le jour 12 à 103,9 euros.
- Vendre le jour 16 à 106,5 euros.
- Acheter le jour 19 à 109,5 euros.

Moyennes mobiles de période 8 :

- Vendre le jour 10 à 101,8 euros.
- Acheter le jour 11 à 102,4 euros.

Le choix de la période dans les moyennes mobiles est déterminant dans le choix. Une période trop courte donne trop de signaux et induit trop de mouvements, alors qu'une période trop longue n'en fournit pas assez. Cette méthode n'est pas parfaite : on constate en effet qu'elle peut inciter dans certains cas à prendre de mauvaises décisions.

OBJECTIFS

- ▶ Savoir calculer des pourcentages d'augmentation ou de diminution
- ▶ Utiliser la moyenne géométrique pour des variations
- ▶ Connaître les indices de Laspeyres, Paasche et Fisher pour un ensemble de produits

PLAN

- 4.1 Mesure d'évolution
- 4.2 Mode de calcul des indices
- 4.3 Exemples d'indices

Les cours boursiers, l'évolution des prix ou encore la production industrielle sont mesurés par des indices. Ces indices informent de manière synthétique. Leur seule donnée indique une progression ou une régression dans un secteur. Nous allons nous intéresser à la manière dont ils se calculent.

4.1 MESURE D'ÉVOLUTION

a) Pourcentage d'évolution

L'année 0, un objet a un prix de 5 euros. L'année 1, son prix passe à 6 euros. L'augmentation de ce prix est donc de 1 euro. Que peut-on en conclure pour un produit d'une valeur de 8 euros l'année 0 ? Rien *a priori*, car si le premier produit représente l'augmentation des prix du marché, son augmentation de 1 euro ne va pas s'appliquer à l'ensemble des prix. Le second produit ne sera pas à 9 euros en année 1.

Considérons l'augmentation **relative** du prix au moyen de la formule :

$$\frac{\text{Prix période 1} - \text{Prix période 0}}{\text{Prix période 0}}$$

L'augmentation **relative** du prix est de 20 %. Cette donnée – si elle représente l'évolution du marché – est utile pour conjecturer l'augmentation de n'importe quel prix.

Le produit à 8 euros passera donc à $8 \times 1,2 = 9,6$ euros.

Avant de réaliser le calcul du prix d'un objet grâce à son taux d'augmentation relative, il faut rappeler que :

Augmentation de t % \iff Multiplication par $(1 + t \text{ %})$

Diminution de t % \iff Multiplication par $(1 - t \text{ %})$

b) Moyenne géométrique

La formule mise en place pour calculer l'augmentation moyenne de données cumulatives est celle de la **moyenne géométrique** :

$$\bar{x} = (x_1 \times x_2 \times \dots \times x_n)^{\frac{1}{n}} = \left(\prod_{i=1}^n x_i \right)^{\frac{1}{n}}$$

La moyenne géométrique s'utilise dans les cas d'effets cumulatifs par exemples les intérêts composés en finance, l'inflation, la croissance...



Considérons l'exemple du taux de croissance annuel du PIB par habitant en volume en France entre 2000 et 2008. Voici les données fournies par l'Insee :

Tableau 4-1 Évolution de la croissance entre 2002 et 2008

Année	2002	2003	2004	2005	2006	2007	2008
Taux de croissance en %	0,3	0,4	1,7	1,1	1,5	1,7	-0,1

On cherche ici à trouver le taux moyen sur la période de début 2002 à fin 2008. Une idée rapide consisterait à calculer une **moyenne arithmétique simple**. Cependant cette idée n'est pas sérieuse ; elle ne donnera pas une véritable idée de moyenne. Pourquoi ?

Il faut prendre en compte l'aspect cumulatif des taux, la croissance de 0,3 % se fait à partir d'une valeur du PIB fin 2001 alors que celle de 2003 se fait à partir d'une valeur du PIB fin 2002.

Déterminons le lien entre le PIB fin 2008 : PIB_{2008} et le PIB fin 2001 : PIB_{2001}

$$PIB_{2002} = 1,003 \times PIB_{2001}$$

$$\text{et } PIB_{2003} = 1,004 \times PIB_{2002} = 1,004 \times 1,003 \times PIB_{2001}$$

$$\text{Ainsi } PIB_{2008} = 0,999 \times 1,017 \times 1,015 \times 1,011 \times 1,017 \\ \times 1,004 \times 1,003 \times PIB_{2001}$$

L'idée d'un taux annuel moyen est celle d'un taux d'une valeur identique chaque année qui permet d'aller du PIB_{2001} au PIB_{2008} .

On cherche donc un taux t vérifiant :

$$PIB_{2008} = (1+t) \times (1+t) \times (1+t) \times (1+t) \times (1+t) \\ \times (1+t) \times (1+t) \times PIB_{2001} = (1+t)^7 \times PIB_{2001}$$

Par identification avec l'équation précédente on trouve :

$$(1+t)^7 = 0,999 \times 1,017 \times 1,015 \times 1,011 \times 1,017 \times 1,004 \\ \times 1,003$$

$$\text{d'où } 1+t = (0,999 \times 1,017 \times 1,015 \times 1,011 \times 1,017 \times 1,004 \\ \times 1,003)^{\frac{1}{7}}$$

et on trouve $t \approx 0,94\%$

Remarque : il serait équivalent pour le PIB d'avoir ce taux moyen chaque année et d'avoir la succession des taux indiqués dans le tableau.

4.2 MODE DE CALCUL DES INDICES

a) Indices élémentaires

L'indice élémentaire d'un prix la période 1 par rapport à une période de référence 0 est donné par la formule suivante en base 100 :

$$I_{1/0} = \frac{P_1}{P_0} \times 100$$

P_1 et P_0 sont les prix à la période 1 et à la période 0. La multiplication par 100 indique que l'indice est en base 100. Il existe des indices en base 1000 comme le CAC40. Ainsi, pour un prix qui passe de 5 euros à

$$6 \text{ euros, on a : } I_{1/0} = \frac{6}{5} \times 100 = 120$$

Les indices élémentaires servent à rendre compte des évolutions relatives des prix en ramenant les prix à 100 à la période 0. Ainsi, ils permet-

tent de comparer facilement des évolutions de prix pour des éléments de valeurs différentes.

Les obligations sont cotées sur les marchés financiers en pourcentage de leur valeur nominale. Ce mode de cotation permet de comparer facilement les obligations entre elles. Supposons qu'une obligation ait un nominal de 500 € et qu'elle cote 495 €, alors la cotation exprimée en pourcentage correspondra à $\frac{495}{500} \times 100 = 99 \%$. Cela exprime le fait que l'obligation a perdu 1 % de sa valeur relativement à sa valeur nominale. Cette caractéristique a facilité le passage à l'euro des obligations car on perd la référence à la valeur nominale.

b) Indice sur plusieurs produits

Les indices sont des indicateurs économiques, pour cela ils doivent rendre compte de l'évolution d'un ensemble de produits.

Exemple. Considérons le tableau suivant :

Tableau 4-2 Évolution des prix de trois produits

	Prix période 0	Prix période 1	Quantité
Produit A	2	3	200
Produit B	5	6	100
Produit C	2	4	50

On peut calculer des indices élémentaires pour chacun des produits du tableau. Cependant, cela ne donne pas une information globale sur les prix. Il faut de plus prendre en compte l'importance en quantité de ces produits. Voyons deux façons de procéder :

Moyenne des indices pondérés

La **moyenne des indices pondérés** consiste à faire une moyenne des indices élémentaires de chaque produit pondéré par les quantités.

$$\text{L'indice élémentaire de A est : } I_{1/0} = \frac{3}{2} \times 100 = 150$$

$$\text{L'indice élémentaire de B est : } I_{1/0} = \frac{6}{5} \times 100 = 120$$

$$\text{L'indice élémentaire de C est : } I_{1/0} = \frac{4}{2} \times 100 = 200$$

Ainsi la **moyenne des indices pondérés** est :

$$I_{1/0} = \frac{150 \times 200 + 120 \times 100 + 200 \times 50}{200 + 100 + 50} = 148,57$$

Cette valeur indique une augmentation des prix de 48,57 %.

Généralisons cet exemple avec la formule de la **moyenne des indices pondérés** :

Notations :

Q_i : quantité associée au produit i

$P_{i,j}$: prix du produit i à la période j

$$I_{1/0} = \frac{Q_1 \frac{P_{1,1}}{P_{1,0}} + Q_2 \frac{P_{2,1}}{P_{2,0}} + \dots + Q_n \frac{P_{n,1}}{P_{n,0}}}{Q_1 + Q_2 + \dots + Q_n} \times 100$$

$$= \frac{\sum_{i=1}^n Q_i \frac{P_{i,1}}{P_{i,0}}}{\sum_{i=1}^n Q_i} \times 100$$

Indice des moyennes pondérées

L'**indice des moyennes pondérées** consiste à globaliser les sommes sur chaque période.

Sur la période 0, la valeur globale est de :

$$2 \times 200 + 5 \times 100 + 2 \times 50 = 1\,000$$

Sur la période 1, la valeur globale est de :

$$3 \times 200 + 6 \times 100 + 4 \times 50 = 1\,400$$

Ainsi, on trouve un indice des moyennes pondérées en base 100 de :

$$\frac{1\,400}{1\,000} \times 100 = 140$$

Cette valeur indique une augmentation globale des prix de 40 %.

Cet exemple est l'illustration de la formule de l'**indice des moyennes pondérées** en base 100 (avec les mêmes notations) :

$$I_{1/0} = \frac{Q_1 P_{1,1} + Q_2 P_{2,1} + \dots + Q_n P_{n,1}}{Q_1 P_{1,0} + Q_2 P_{2,0} + \dots + Q_n P_{n,0}} \times 100$$

$$= \frac{\sum_{i=1}^n Q_i P_{i,1}}{\sum_{i=1}^n Q_i P_{i,0}} \times 100$$

Remarque : on peut constater que les deux formules fournissent des résultats sensiblement différents. Il n'est pas équivalent de constater une augmentation des prix de 40 % et de 48,57 %. En pratique, les deux formules sont utilisées d'où une incertitude sur la pertinence du résultat.

c) Quantités et prix

L'exemple précédent ne prend pas en compte la variation éventuelle des quantités entre la période 0 et la période 1.

Exemple. Intéressons nous à l'exemple suivant :

Tableau 4-3 Évolution des prix et des quantités de trois produits

	Prix période 0	Prix période 1	Quantité période 0	Quantité période 1
Produit A	2	2,5	150	100
Produit B	5	6	100	80
Produit C	2	3	50	45

Il se pose la question de ce que l'on cherche à mesurer : la variation des quantités ou la variation des prix ?



Erreur à éviter : chercher à mesurer les deux en même temps avec une formule du type : $\frac{2,5 \times 100 + 6 \times 80 + 3 \times 45}{2 \times 150 + 5 \times 100 + 2 \times 50} \times 100 = 96,11$. Comment interpréter ce résultat ? Indique-t-il une variation de prix ou de quantités ? Ce résultat n'a en réalité aucun sens car on ne peut mesurer la variation spécifique d'un paramètre en faisant varier les deux en même temps. Il faut donc choisir de fixer les prix quand on cherche un indice des quantités et il faut fixer les quantités quand on cherche un indice des prix.

À quelle valeur fixer les prix ou les quantités ? Il y a deux façons de fixer ces paramètres : soit à la période de référence (période 0), on parle alors d'indices de **Laspeyres** ; soit à la période actuelle (période 1), on parle alors d'indices de **Paasche**.

Voyons les différents calculs possibles de l'**indice des moyennes pondérées** :

- Indice de Laspeyres des prix :

$$\frac{2,5 \times 150 + 6 \times 100 + 3 \times 50}{2 \times 150 + 5 \times 100 + 2 \times 50} \times 100 = 125$$

Soit une augmentation des prix de 25 %.

- Indice de Paasche des prix :

$$\frac{2,5 \times 100 + 6 \times 80 + 3 \times 45}{2 \times 100 + 5 \times 80 + 2 \times 45} \times 100 = 125,36$$

Soit une augmentation des prix de 25,36 %.

- Indice de Laspeyres des quantités :

$$\frac{2 \times 100 + 5 \times 80 + 2 \times 45}{2 \times 150 + 5 \times 100 + 2 \times 50} \times 100 = 76,67$$

Soit une baisse de 23,33 %.

- Indice de Paasche des quantités :

$$\frac{2,5 \times 100 + 6 \times 80 + 3 \times 45}{2,5 \times 150 + 6 \times 100 + 3 \times 50} \times 100 = 76,89$$

Soit une baisse de 23,11 %.

Résumons les formules utiles à ces calculs ainsi :

Notation : $Q_{i,j}$: quantité du produit i à la période j et $P_{i,j}$: prix du produit i à la période j .

Tableau 4-4 Formules de Paasche et de Laspeyres

	Indice des quantités	Indice des prix
Indice de Laspeyres	$\frac{\sum_{i=1}^n Q_{i,1} P_{i,0}}{\sum_{i=1}^n Q_{i,0} P_{i,0}} \times 100$	$\frac{\sum_{i=1}^n Q_{i,0} P_{i,1}}{\sum_{i=1}^n Q_{i,0} P_{i,0}} \times 100$
Indice de Paasche	$\frac{\sum_{i=1}^n Q_{i,1} P_{i,1}}{\sum_{i=1}^n Q_{i,0} P_{i,1}} \times 100$	$\frac{\sum_{i=1}^n Q_{i,1} P_{i,1}}{\sum_{i=1}^n Q_{i,1} P_{i,0}} \times 100$

Il apparaît que les valeurs données par l'indice de **Paasche** et l'indice de **Laspeyres** sont proches. Pour ne pas avoir à choisir entre les deux, ou plutôt pour prendre en compte les deux, l'indice de **Fisher** est utilisé. L'indice de **Fisher** est la moyenne géométrique des deux indices précédents :

$$\text{Indice de Fisher} = \sqrt{\text{Indice de Laspeyres} \times \text{Indice de Paasche}}$$

4.3 EXEMPLES D'INDICES

- **L'indice des prix à la consommation (IPC)** mesure le niveau moyen des prix des biens et services consommés par les ménages. Il se calcule sur un ensemble de 200 000 prix par la méthode de Laspeyres avec l'année 2000 pour référence. Il permet de mesurer l'inflation.
- **L'indice de la production industrielle (IPI)** mesure les variations des quantités produites du secteur secondaire, l'industrie, y compris des industries agricoles et alimentaires, de l'énergie et de la construction. Il s'agit d'un indice de Laspeyres des quantités, base 100 en 2005.
- **Les indices annuels de la construction** dans l'Union européenne, sont des indices Laspeyres de base 100 en 2005 qui évaluent dans chaque pays les évolutions des entrées de commande, des permis de construire et du prix des nouveaux bâtiments résidentiels.
- **L'indice des prix de gros alimentaire** mesure l'évolution des prix payés par les détaillants de la région parisienne. Il s'agit d'un indice de Laspeyres de base 100 en 2005.
- **Le CAC40** est l'indice de la bourse de Paris, il est composé des 40 plus grandes capitalisations à la bourse de Paris. Il s'agit d'un indice de base 1 000 créé le 31 décembre 1987. L'inflation est prise en compte ainsi que la quantité de titres réellement sur les marchés (depuis 2003).
- **Le Dow Jones et le Nikkei 225** sont les indices des bourses de New-York et de Tokyo. Ils ont la particularité d'être pondérés sur la valeur des actions les composant et non sur leur capitalisation boursière. Cela explique des difficultés d'utilisation de ces indices.



POINTS CLEFS

► Mesure de variations :

Variation	Opération
+10 %	$\times 1,1$
-10 %	$\times 0,9$
+1 %	$\times 1,01$
-1 %	$\times 0,99$
+t %	$\times (1 + t \%)$
-t %	$\times (1 - t \%)$

- **Moyenne géométrique** : dans le cas d'une suite de variations, le coefficient de variation moyen se calcule à l'aide de la moyenne géométrique en faisant le produit des coefficients. On élève ensuite ce produit à la puissance $\frac{1}{\text{nombre de coefficients}}$. À l'aide du tableau précédent, on retrouve le pourcentage de variation moyen.

- **Indice élémentaire d'un prix entre deux périodes**

$$I_{1/0} = \frac{P_1}{P_0} \times 100$$

- **Indices sur plusieurs produits**

- On distingue la *moyenne des indices pondérés* de l'*indice des moyennes pondérées*, lorsque la quantité est fixe.
- Si la quantité varie, on retient l'*indice des moyennes pondérées* et on se réfère à une donnée actuelle pour l'*indice de Paasche* et à une donnée de référence passée pour l'*indice de Laspeyres*.
- On calcule soit des indices de quantités, soit des indices de prix.
- L'*indice de Fisher* est la moyenne géométrique des deux indices précédents.

EXERCICES

4.1 Moyenne

Un objet voit son prix augmenter de 3 % par ans pendant 5 ans, puis de 60 % en 10 ans et passe ensuite en 5 ans de l'indice 135 à l'indice 178, pour baisser ensuite de 2 % par ans pendant 4 ans.

Quel est le taux moyen annuel d'évolution de ce prix ?

4.2 Évolution de prix

On s'intéresse à l'évolution du prix de deux produits sur plusieurs années :

	2004	2006	2008	2010
Produit A	2,1	2,3	2,6	2,6
Produit B	12,4	12,6	13	13,5

- a) Pour comparer ces évolutions, on choisit de considérer l'année 2004 comme référence. Donner les indices élémentaires des deux produits pour les années 2006, 2008 et 2010. Quel produit a subi la plus forte progression ?
- b) En déduire le pourcentage d'augmentation annuel moyen pour chacun de ces deux produits.

4.3 Immobilier

On achète un bien immobilier de 100 000 euros dans trois pays différents. Dix ans plus tard, on souhaite revendre ces biens. Le prix de revente dans le premier pays est de 95 000 euros, dans le second de 100 000 euros, et dans le dernier de 120 000 euros. Sachant qu'entre l'achat et la vente, l'inflation a été de -10% dans le premier pays, de 15% dans le second et de 20% dans le dernier. Quelle opération vous semble la plus intéressante ?

4.4 Investissement

Une entreprise achète en 2010 un appareil dont le prix est 9 700 euros. Le même modèle avait pour valeur 8 900 euros en 2006.

- a) Déterminer $I_{2010/2006}$ l'indice élémentaire de ce prix en base 100 en 2006.
- b) En déduire la valeur du même modèle en 2014 en supposant que $I_{2010/2006} = I_{2014/2010}$

4.5 Paasche, Laspeyres et Fisher

Calculer l'indice de Laspeyres, de Paasche et de Fisher pour le nombre d'appartements à acheter et pour les prix du m^2 en euros de l'immobilier à Paris :

	Prix du m^2 en 2000	Prix du m^2 en 2010	À acheter en 2000	À acheter en 2010
Quartier A	2 100	5 030	1 500	1 000
Quartier B	3 900	10 910	800	990
Quartier C	2 800	6 820	5 000	4 400

SOLUTIONS

Exercice 4.1

Notons P le prix initial, il subit une augmentation de 3 % par ans pendant 5 ans. Ainsi dans 5 ans son prix sera de $1,03^5 P$. Ensuite, en dix ans, il augmente de 60 %. Dans 15 ans, son prix sera donc de $1,6 \times 1,03^5 P$. Le prix passe ensuite de l'indice 135 à 178 en 5 ans. Il s'agit d'une augmentation de $\frac{178 - 135}{135} = 31,85 \%$. Le prix dans 20 ans est donc de $\frac{178}{135} \times 1,6 \times 1,03^5 P$. Il baisse encore de 2 % par ans pendant 4 ans. Le prix passe donc à $0,98^4 \times \frac{178}{135} \times 1,6 \times 1,03^5 P$ en 24 ans. Le taux moyen annuel est donné par :

$$(0,98^4 \times \frac{178}{135} \times 1,6 \times 1,03^5)^{\frac{1}{24}} = 1,03448$$

Soit une augmentation moyenne de 3,45 % par an.

Exercice 4.2

a) Présentons sous forme de tableau les indices élémentaires correspondant aux deux produits pour comparer les évolutions :

	2004	2006	2008	2010
Produit A	100	109,53	123,81	123,81
Produit B	100	101,61	104,84	108,87

On constate ainsi que le produit A augmente de 23,81 % sur la période alors que le produit B augmente seulement de 8,87 %.

b) Entre 2004 et 2010, il y a 6 ans.

L'augmentation moyenne de A est donc de $1,2381^{1/6} - 1 = 3,51 \%$ par an. Celle de B de $1,0887^{1/6} - 1 = 1,43 \%$ par an.

Exercice 4.3

Sans inflation, la comparaison est évidente. Avec l'inflation, il s'agit de déterminer la valeur réelle des biens immobiliers au moment de la revente. Dans le premier pays, l'inflation est de -10 %. Une valeur de 100 000 euros à l'achat devrait correspondre à une valeur de 90 000 euros à la vente si la valeur du bien ne fait que suivre l'inflation. Ici, le bien a une valeur de revente de 95 000 euros. Le bien a donc pris de la valeur en dix ans, malgré l'impression d'une baisse. On a réalisé

une plus-value de $\frac{95\,000 - 90\,000}{90\,000} = 5,56\%$. Dans le second pays, le bien ne change pas de prix ; pourtant l'inflation est de 15 %. Ainsi, il devrait avoir une valeur de revente de 115 000 euros. Il y a donc une perte de valeur : $\frac{100\,000 - 115\,000}{115\,000} = -13,04\%$. Dans le dernier pays le prix de revente est de 120 000 euros. L'inflation étant de 20 %, le bien n'a pas changé de valeur. L'opération la plus intéressante est donc celle du premier pays.

Exercice 4.4

$$\text{a) On a } I_{2010/2006} = \frac{9\,700}{8\,900} \times 100 = 108,99$$

$$\text{b) On a } I_{2014/2010} = \frac{\text{Prix en 2014}}{\text{Prix en 2010}} \times 100$$

$$\text{Or } I_{2014/2010} = I_{2010/2006} = 108,99$$

$$\text{Donc } \frac{\text{Prix en 2014}}{9\,700} \times 100 = 108,99$$

$$\text{Ainsi : Prix en 2014} = \frac{108,99 \times 9\,700}{100} = 10\,571,91 \text{ euros}$$

Exercice 4.5

Résultats sous forme de tableau :

	Indice des prix	Indice des quantités
Indice de Laspeyres	$\frac{5\,030 \times 1\,500 + 10\,910 \times 800 + 6\,820 \times 5\,000}{2\,100 \times 1\,500 + 3\,900 \times 800 + 2\,800 \times 5\,000} \times 100$ = 248,51	$\frac{2\,100 \times 1\,000 + 3\,900 \times 990 + 2\,800 \times 4\,400}{2\,100 \times 1\,500 + 3\,900 \times 800 + 2\,800 \times 5\,000} \times 100$ = 90,19
Indice de Paasche	$\frac{5\,030 \times 1\,000 + 10\,910 \times 990 + 6\,820 \times 4\,400}{2\,100 \times 1\,000 + 3\,900 \times 990 + 2\,800 \times 4\,400} \times 100$ = 250,75	$\frac{5\,030 \times 1\,000 + 10\,910 \times 990 + 6\,820 \times 4\,400}{5\,030 \times 1\,500 + 10\,910 \times 800 + 6\,820 \times 5\,000} \times 100$ = 91,00
Indice de Fishe	$\sqrt{248,51 \times 250,75} = 249,63$	$\sqrt{90,19 \times 91,00} = 90,59$

PARTIE 2

Probabilités

Chapitre 5	Notions de bases de probabilités	79
Chapitre 6	Lois fondamentales de probabilités	109
Chapitre 7	Estimateurs et tests d'hypothèses	141
Chapitre 8	Test du χ^2	161

Au-delà de la multiplicité des paramètres, les phénomènes économiques souffrent d'une part d'aléatoire. La modélisation probabiliste permet de prendre en compte l'aléa d'un phénomène pour réaliser une prévision vraisemblable.

Les probabilités servent au gestionnaire pour proposer une réponse à la demande d'un marché, pour constituer un stock de sécurité ou encore pour estimer un risque de défaillance dans la production.

La seconde partie de l'ouvrage se consacre à la construction de modèles probabilistes appliqués à la gestion.

Le cinquième chapitre, « Notions de bases de probabilités », a pour but de mettre en place les outils essentiels à la modélisation d'un phénomène aléatoire. Le sixième chapitre, « Lois fondamentales de probabilités », présente les lois les plus couramment utilisées par le gestionnaire. Le septième chapitre, « Estimateurs et tests d'hypothèses », propose la modélisation de phénomènes généraux à partir de la connaissance d'un échantillon. Le huitième chapitre, « Test du χ^2 », permet d'apprécier la validité d'une modélisation ou d'une relation.

CHAPITRE 5

Notions de base de probabilités

OBJECTIFS

- ▶ Maîtriser les éléments du dénombrement (p-listes, arrangements et combinaisons)
- ▶ Savoir calculer la probabilité d'un événement
- ▶ Connaître les propriétés essentielles du calcul des probabilités (probabilités conditionnelles, événements complémentaires, indépendance...)
- ▶ Donner la loi d'une variable aléatoire discrète ou continue
- ▶ Calculer l'espérance et l'écart-type d'une variable aléatoire
- ▶ Représenter graphiquement une fonction de répartition

PLAN

- 5.1 Dénombrement
- 5.2 Calcul de probabilités
- 5.3 Lois de probabilités

Qu'est-ce que comprendre un phénomène ? Par exemple, comment comprendre le comportement des clients dans un magasin ? Comment comprendre l'arrivée de clients dans un centre d'appels téléphoniques ?

Dans un premier temps, des données chiffrées peuvent être observées et compulsées. Les outils **statistiques** de moyenne, d'écart-type et de quartiles donnent une synthèse de nos observations. Les outils de régression permettent, avec un certain risque, d'envisager des prévisions.

Ces outils suffisent-ils vraiment à comprendre un phénomène ? Les données statistiques de base ne servent qu'à faire une synthèse des données. Les modèles de régression servent à faire des prévisions pour prolonger un ensemble d'observations passées. Que faire lors du lancement d'un nouveau produit pour lequel on ne dispose d'aucunes observations ? On

a besoin de modèles qui englobent un ensemble de situations équivalentes. Les **probabilités** proposent une abstraction supplémentaire par rapport aux **statistiques**. Elles envisagent des modèles détachés d'un cas particulier d'observation qui permettent de faire des prévisions sur des familles plus larges d'expériences.

Ce chapitre est une présentation de la théorie de base des probabilités avec des illustrations en finance et en économie.

5.1 DÉNOMBREMENT

L'outil du dénombrement est indispensable pour le calcul des probabilités dans les cas finis discrets. On entend par « finis discrets » des cas pour lesquels on peut compter (dénombrer) les issues possibles. Par exemple, un lancé de dé est une expérience dont les résultats sont dans l'ensemble $\{1,2,3,4,5,6\}$. Un contre-exemple est le temps d'attente à une caisse, on ne peut *a priori* pas lister l'ensemble des résultats possibles.

Cette partie est une présentation des éléments les plus classiques du dénombrement.

a) p-listes

Exemple. Une banque distribue à chacun de ses clients un identifiant de connexion formé de six chiffres entre 0 et 9. Combien cette banque peut-elle avoir de clients au maximum avec ce mode de connexion ?

Il est clair que deux clients ne peuvent pas avoir le même identifiant et que chaque client a un unique identifiant. L'ordre des chiffres a une importance mais chaque chiffre peut être utilisé plusieurs fois. Ainsi, il y a dix possibilités pour le premier chiffre de l'identifiant, il y en a aussi dix pour le second, et à chaque fois dix chiffres possibles pour chaque chiffre de l'identifiant. Ainsi on a $10 \times 10 \times 10 \times 10 \times 10 \times 10 = 10^6 = 1\,000\,000$ identifiants possibles, par conséquent la banque peut avoir au plus 1 000 000 de clients avec ce mode de connexion.

Cet exemple est l'illustration de la formule de **dénombrement des p-listes** :

n^p où n représente le nombre d'éléments possible et p la longueur de la liste.



Dans quel cas utiliser cette formule ?

Cette formule s'utilise quand l'expérience s'apparente à un **tirage successif avec remise**. Dans l'exemple, les éléments successifs à tirer sont des chiffres de 0 à 9, il y a remise car chaque élément peut être tiré plusieurs fois.

b) Arrangements

Considérons l'exemple précédent avec la contrainte de n'utiliser qu'une seule fois chaque chiffre, pour le premier chiffre du code, il y a 10 choix possibles. Pour le second, il n'y en a plus que 9 : tous les chiffres de 0 à 9 sauf celui utilisé en premier. Pour le troisième, il n'y en a plus que 8. Ainsi le nombre de codes possible est $10 \times 9 \times 8 \times 7 \times 6 \times 5 = 151\,200$. Le nombre de clients possible est alors 151 200.

Cet exemple est l'illustration de la formule de **dénombrement des arrangements** :

$$n \times (n - 1) \times (n - 2) \times (n - 3) \times \dots \times (n - p + 1) = \frac{n!}{(n - p)!} = A_n^p$$

où n représente le nombre d'éléments possible et p la longueur de l'arrangement.

Rappel : $n! = 1 \times 2 \times 3 \times 4 \times \dots \times n$



Dans quel cas utiliser cette formule ?

Cette formule s'utilise quand l'expérience s'apparente à un **tirage successif sans remise**. Dans l'exemple, les éléments successifs à tirer sont des chiffres de 0 à 9, il n'y a pas remise car chaque élément ne peut être tiré qu'une fois.

c) Combinaisons

Exemple. On doit choisir deux salariés dans une équipe de cinq salariés pour effectuer une mission spécifique. De combien de façon peut-on choisir ces deux salariés ?

A priori tous les couples sont possibles, en notant les salariés A, B, C, D et E les différents couples possibles sont (A,B) ; (A,C) ; (A,D) ; (A,E) ;

(B,C) ; (B,D) ; (B,E) ; (C,D) ; (C,E) et (D,E) soit en tout dix couples possibles. On remarque que l'ordre n'a pas d'importance : le couple (A,B) est le même que (B,A).

Dénombrement des combinaisons

Une manière plus rapide de déterminer le nombre de couples possible est d'utiliser la formule de **dénombrement des combinaisons** :

$$\binom{n}{p} = \frac{n!}{p!(n-p)!}$$

où n représente le nombre d'éléments et p la taille de la combinaison.



Dans quel cas utiliser cette formule ?

Cette formule s'utilise quand l'expérience s'apparente à un **tirage simultané**. Il n'y a pas d'ordre dans le tirage. Dans l'exemple, les deux salariés sont sélectionnés en même temps.

Comment utiliser cette formule ?

L'utilisation des factorielles entraîne des calculs longs. Il est préférable de réaliser des simplifications avant calcul :

Exemple.

$$\bullet \binom{5}{2} = \frac{5!}{2!(5-2)!} = \frac{5 \times 4 \times \cancel{3} \times \cancel{2} \times \cancel{1}}{2 \times 1 \times \cancel{3} \times \cancel{2} \times \cancel{1}} = \frac{5 \times 4}{2} = 10$$

$$\begin{aligned} \bullet \binom{39}{3 \ 3} &= \frac{39!}{33!(39-33)!} \\ &= \frac{39 \times 38 \times 37 \times 36 \times 35 \times 34 \times \cancel{33} \times \cancel{32} \dots \times \cancel{1}}{\cancel{33} \times \cancel{32} \dots \times \cancel{1} \times 6 \times 5 \times 4 \times 3 \times 2 \times 1} \\ &= \frac{39 \times 38 \times 37 \times 36 \times 35 \times 34}{6 \times 5 \times 4 \times 3 \times 2 \times 1} = 3\,262\,623 \end{aligned}$$

Propriétés

Certaines **propriétés** évitent de perdre du temps en calculs :

$$1. \binom{n}{0} = \binom{n}{n} = 1, \text{ ainsi } \binom{23}{23} = 1$$

$$2. \binom{n}{1} = \binom{n}{n-1} = n, \text{ ainsi } \binom{39}{38} = 39$$

$$3. \binom{n}{p} = \binom{n}{n-p}, \text{ ainsi } \binom{5}{2} = \binom{5}{3} = 10$$

Le triangle de Pascal

Le **triangle de Pascal** permet de calculer simplement les premières valeurs de $\binom{n}{p}$ grâce à la relation : $\binom{n-1}{p-1} + \binom{n-1}{p} = \binom{n}{p}$

Tableau 5-1 Triangle de Pascal

	0	1	2	3	4	5	6	7
0	1							
1	1	1						
2	1	2	1					
3	1	3	3	1				
4	1	4	6	4	1			
5	1	5	10	10	5	1		
6	1	6	15	20	15	6	1	
7	1	7	21	35	35	21	7	1

Notons que les chiffres en gras dans la première ligne sont les valeurs de p et les chiffres en gras dans le première colonne sont les valeurs de n .

Mode de construction du tableau

– Le premier et le dernier élément d'une ligne est toujours 1, car :

$$\binom{n}{0} = \binom{n}{n} = 1$$

– La relation $\binom{n-1}{p-1} + \binom{n-1}{p} = \binom{n}{p}$ permet de compléter le tableau. Par exemple, on a $\binom{3}{1} + \binom{3}{2} = 3 + 3 = 6$, donc $\binom{4}{2} = 6$.



5.2 CALCUL DE PROBABILITÉS

a) Comment définir une probabilité ?

Exemple. Un banquier propose un livret A à ses clients. Il mesure la fréquence de sa réussite :

$$\frac{\text{Nombre de livrets A ouverts}}{\text{Nombre de propositions}} = \frac{n_A}{n}$$

Combien de propositions le banquier doit-il faire pour évaluer sa fréquence moyenne de réussite ? Comment passer d'une **observation** à une **conceptualisation** du phénomène ? La notion de **probabilité** permet de faire ce passage. Le nombre d'observations nécessaires doit être infini pour passer de la mesure de fréquence au concept de probabilité.

On notera R l'événement : « Proposition ayant aboutie à l'ouverture d'un livret A ».

La probabilité de cet événement se note $P(R)$ et se définit par :

$$P(R) = \lim_{n \rightarrow \infty} \frac{n_A}{n}$$

Cette conceptualisation est le passage des **statistiques** à la **théorie des probabilités**.

b) Propriétés des probabilités

1. Une probabilité étant une limite de fréquence, les valeurs possibles pour une probabilité sont dans l'intervalle $[0 ; 1]$.
2. On note Ω l'ensemble des possibles ou l'événement certain et \emptyset l'ensemble vide ou événement impossible. On a $P(\Omega) = 1$ et $P(\emptyset) = 0$

Remarque : un événement est une partie de l'ensemble des possibles. Par exemple dans un lancer de dé, l'événement « obtenir un nombre pair » est une partie de l'ensemble des résultats possibles : $\{ 1 ; 2 ; 3 ; 4 ; 5 ; 6 \}$

3. Soient A et B deux événements. On note $A \cap B$ l'intersection des événements A et B , autrement dit l'ensemble des éléments à la fois dans A et dans B . On note $A \cup B$ l'union des événements A et B , autrement dit l'ensemble des éléments dans A ou dans B .

On représente schématiquement ces événements ainsi :

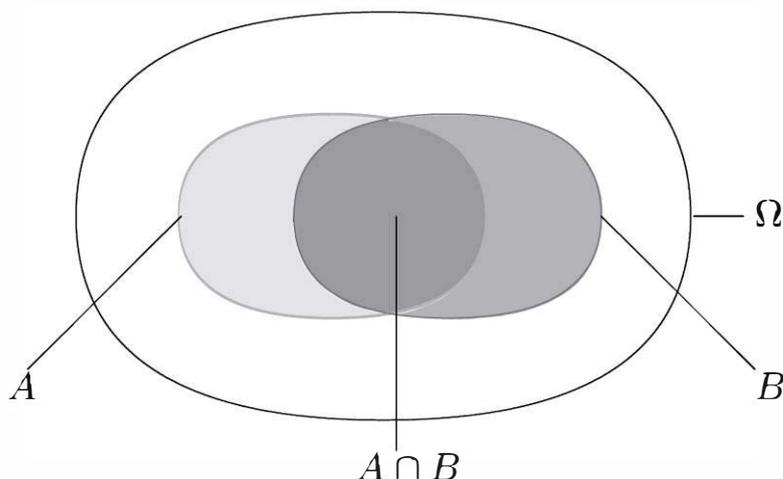


Figure 5-1 Représentation schématique des événements

On a la relation suivante :

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Si A et B sont **incompatibles**, autrement dit si $A \cap B = \emptyset$, on a :

$$P(A \cup B) = P(A) + P(B)$$

Exemple. Deux questions sont posées dans un questionnaire. 40 % des personnes répondent oui aux deux questions, 56 % répondent oui à la première question et 78 % répondent oui à l'une ou l'autre des deux questions. Quel pourcentage a répondu oui à la deuxième question ?

Il ne s'agit pas explicitement de probabilités mais le raisonnement est identique. Notons O_1 l'événement « répondre oui à la première question » et O_2 l'événement « répondre oui à la deuxième question ».

On a $P(O_1) = 0,56$ et $P(O_1 \cap O_2) = 0,4$ et $P(O_1 \cup O_2) = 0,78$

D'après la relation précédente :

$$P(O_1 \cup O_2) = P(O_1) + P(O_2) - P(O_1 \cap O_2)$$

ainsi $0,78 = 0,56 + P(O_2) - 0,4$ soit $P(O_2) = 0,62$

4. On note \bar{A} le contraire de l'événement A . Il se définit par :

$$\bar{A} \cup A = \Omega$$

$$\bar{A} \cap A = \emptyset$$

On a donc avec la relation précédente :

$$P(\bar{A} \cup A) = P(A) + P(\bar{A}) - P(\bar{A} \cap A)$$

d'où $P(\Omega) = P(A) + P(\bar{A}) - P(\emptyset)$, soit $1 = P(A) + P(\bar{A})$, ainsi :

$$P(\bar{A}) = 1 - P(A)$$

Quel pourcentage a répondu non à la question 2 ?

Répondre non à la deuxième question est le contraire de répondre oui à la deuxième question.

$$\text{Ainsi } P(\bar{O}_2) = 1 - P(O_2) = 1 - 0,62 = 0,38$$

Remarque : il arrive qu'un événement soit difficile à déterminer directement. L'événement contraire peut s'avérer plus simple à déterminer.

5. Le contraire logique de l'union est l'intersection ainsi :

$$\bullet \overline{A \cup B} = \bar{A} \cap \bar{B}, \text{ ainsi } P(\overline{A \cup B}) = P(\bar{A} \cap \bar{B})$$

$$\bullet \overline{A \cap B} = \bar{A} \cup \bar{B}, \text{ ainsi } P(\overline{A \cap B}) = P(\bar{A} \cup \bar{B})$$

Quel pourcentage a répondu non aux deux questions ?

On s'intéresse à l'événement $\bar{O}_1 \cap \bar{O}_2$ qui est aussi l'événement $\overline{O_1 \cup O_2}$. Or $P(O_1 \cup O_2) = 0,78$

$$\text{d'où } P(\bar{O}_1 \cap \bar{O}_2) = 1 - 0,78 = 0,22$$

c) Le cas d'équiprobabilité

Exemple. Une usine a produit 346 pièces dont 23 sont défectueuses. On prélève au hasard une pièce, quelle est la probabilité qu'elle soit défectueuse ?

$$\text{La réponse est évidente : } \frac{23}{346} = 6,65 \%$$

Nous sommes dans un cas **d'équiprobabilité**, chaque événement est constitué d'éléments qui ont chacun la même probabilité d'apparition. Dans l'exemple les éléments sont les pièces. Ces pièces ont chacune une probabilité $\frac{1}{346}$ d'être prélevée. L'événement « être défectueuse » est réalisé par 23 pièces sur 346.

Cet exemple très simple est une illustration de la **formule d'équiprobabilité** :

$$P = \frac{\text{Nombre de cas favorables}}{\text{Nombre de cas possibles}}$$

Remarque : dans la pratique, le dénombrement du nombre de cas favorables n'est pas toujours simple, les arrangements et les combinaisons seront des outils utiles pour ces calculs.

d) Probabilités conditionnelles

Dépendance d'événements

Exemple. Deux usines produisent des pièces. La production de l'usine A est défectueuse à 5 %. La production de l'usine B est défectueuse à 3 %. L'usine A réalise 40 % de la production totale. Quel pourcentage de la production est défectueux pour l'ensemble des deux usines ? Si une pièce est défectueuse, quelle est la probabilité qu'elle provienne de l'usine A ?

Pour répondre à ces questions, on envisage de schématiser le problème par un **arbre** :

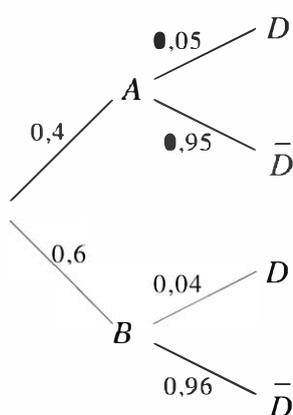


Figure 5-2 Arbre de probabilités

Notation : D : « Être défectueux »

A : « Provenir de l'usine A »

B : « Provenir de l'usine B »

Méthode de calcul

- Sur une même branche, les probabilités se multiplient.
- Sur deux branches différentes, les probabilités s'ajoutent.

On trouve ainsi $P(D) = 0,4 \times 0,05 + 0,6 \times 0,04 = 4,4 \%$.

Remarque : la schématisation par un arbre évite de formaliser le problème mais ne permet pas de répondre à la seconde question de l'exemple.

Pour cela, introduisons la **formule des probabilités conditionnelles**.

La probabilité de A sachant B est donnée par la formule :

$$P_B(A) = \frac{P(A \cap B)}{P(B)}.$$

La seconde question est la recherche de la probabilité : $P_D(A)$. D'après la formule précédente : $P_D(A) = \frac{P(A \cap D)}{P(D)}$.

On a déjà calculé $P(D) = 4,4\%$ et d'après l'arbre précédent $P(A \cap D) = 0,4 \times 0,05 = 2\%$.

Ainsi $P_D(A) = \frac{2\%}{4,4\%} = 45,45\%$.

Indépendance

Deux événements ne sont pas toujours dépendants. On dit que A et B sont **indépendant** si $P_A(B) = P(B)$ ou $P_B(A) = P(A)$. Cela indique que la réalisation de B n'a aucune incidence sur celle de A et réciproquement.

Une propriété est souvent utile pour montrer que deux événements sont indépendants :

$$A \text{ et } B \text{ sont indépendants} \iff P(A \cap B) = P(A) \times P(B)$$

Remarque : il s'agit d'une équivalence. La propriété $P(A \cap B) = P(A) \times P(B)$ prouve que les événements sont indépendants et si les événements sont indépendants alors :

$$P(A \cap B) = P(A) \times P(B)$$

e) Partition d'événements

On dit que la famille d'événements $E_1, E_2, E_3, \dots, E_n$ forme une partition de l'ensemble des possibles si :

- $E_i \cap E_j = \emptyset$ si $i \neq j$
- $E_1 \cup E_2 \cup E_3 \dots E_n = \Omega$

Une partition d'événements peut servir à calculer la probabilité d'un événement A par la formule :

$$P(A) = P(A \cap E_1) + P(A \cap E_2) + \dots + P(A \cap E_n)$$

Remarque : deux événements complémentaires forment une partition de l'ensemble des possibles. En effet, si A est un événement et \bar{A} son complémentaire on a :

- $A \cap \bar{A} = \emptyset$
- $A \cup \bar{A} = \Omega$

Ainsi, si B est un autre événement, en adaptant la formule précédente avec $E_1 = A$ et $E_2 = \bar{A}$, on trouve $P(B) = P(B \cap A) + P(B \cap \bar{A})$.

Cette formule est plus utilisée que la formule générale précédente. On dispose rarement d'une famille d'événements E_1, E_2, \dots, E_n formant une partition de l'ensemble des possibles. Par contre, si on dispose d'un seul événement A la connaissance de son complémentaire suffit pour appliquer la formule précédente.

Par exemple, si on connaît la probabilité de l'événement A : « Le CAC 40 est en baisse » et que l'on cherche la probabilité de l'événement B : « Le cours de l'action EDF monte », on peut déterminer la probabilité que le cours de l'action EDF monte et que le CAC40 soit en baisse : $P(A \cap B)$, puis la probabilité que le cours de l'action EDF monte et que le CAC40 ne soit pas en baisse : $P(\bar{A} \cap B)$. On trouve ainsi la probabilité que le cours de l'action EDF monte : $P(B)$ en additionnant les deux probabilités précédentes.

5.3 LOIS DE PROBABILITÉS

a) Variable discrète

Loi de probabilité

Exemple. Un service de recrutement pose deux questions à ses postulants. Un tiers d'entre eux répond faux aux deux questions. Un tiers d'entre eux répond juste à une des deux questions. Un tiers d'entre eux répond juste aux deux questions. Il reste deux candidats en concurrence pour le poste. Quels sont les écarts de résultats possibles ? Quelle est la probabilité de chaque écart ?

On s'intéresse ici à une **variable aléatoire** : l'écart, que l'on notera E . On cherche à déterminer l'ensemble des valeurs qui peuvent être prises par celle-ci. Le tableau suivant permet de visualiser les différents écarts possibles entre les candidats. La première ligne en gras représente le nombre de bonnes réponses du premier candidat et la première colonne en gras représente le nombre de bonnes réponses du second candidat.

Tableau 5-2 Ensemble des possibles pour E

	0	1	2
0	0	1	2
1	1	0	1
2	2	1	0

L'ensemble des possibles est donc $\{0 ; 1 ; 2 \}$.

On peut calculer pour chaque valeur de E sa probabilité associée. Cette association définit la **loi de probabilité** de E . Chaque case du tableau précédent est équiprobable car chaque candidat répond juste avec la même probabilité à 0, 1 ou 2 questions. Ainsi, à l'aide d'un dénombrement simple, on trouve :

$$P(E = 0) = \frac{3}{9}$$

$$P(E = 1) = \frac{4}{9}$$

$$P(E = 2) = \frac{2}{9}$$

Tableau 5-3 Loi de probabilité de E

e_i	0	1	2
p_i	$\frac{3}{9}$	$\frac{4}{9}$	$\frac{2}{9}$

Remarque : la somme des probabilités est égale à 1.

Formalisation :

La **loi de probabilité** d'une variable aléatoire X est l'association $(x_i; p_i)$ où x_i est une valeur possible de X et $p_i = P(X = x_i)$.

Espérance et variance

Espérance : la notion d'espérance en probabilités rejoint la notion de moyenne en statistiques. L'espérance est un indicateur de tendance centrale pour la variable aléatoire.

Formule de l'espérance :

$$E(X) = \sum_{i \in \Omega} x_i \times p_i$$

Dans l'exemple précédent, on trouve :

$$E(\text{écart}) = 0 \times \frac{3}{9} + 1 \times \frac{4}{9} + 2 \times \frac{2}{9} = \frac{8}{9} \approx 0,89$$

Écart-type : comme en statistiques, l'écart-type est un indicateur de dispersion. L'écart-type est la racine carrée de la variance.

Formule de la variance :

$$V(X) = E(X^2) - (E(X))^2 = \sum_{i \in \Omega} x_i^2 \times p_i - \left(\sum_{i \in \Omega} x_i \times p_i \right)^2$$

Ainsi la **formule de l'écart-type** est :

$$\sigma(X) = \sqrt{V(X)} = \sqrt{E(X^2) - (E(X))^2}$$

Dans l'exemple précédent, on trouve :

$$V(\text{écart}) = 0^2 \times \frac{3}{9} + 1^2 \times \frac{4}{9} + 2^2 \times \frac{2}{9} - \left(\frac{8}{9} \right)^2 = \frac{44}{81}$$

$$\text{Ainsi } \sigma(\text{écart}) = \sqrt{\frac{44}{81}} \approx 0,737$$

Distribution et fonction de répartition

On s'intéresse dans cette partie aux modes de représentation d'une variable aléatoire. On prendra comme exemple la variable aléatoire E .

On représente la **distribution d'une variable aléatoire** discrète par un diagramme en bâton. Chaque bâton représente la probabilité $P(X = x_i)$. L'interprétation de ce diagramme est très simple.

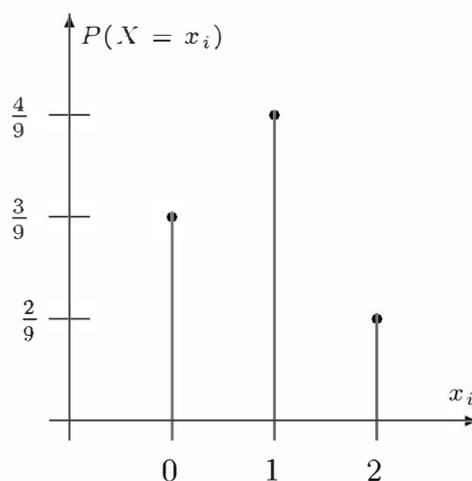
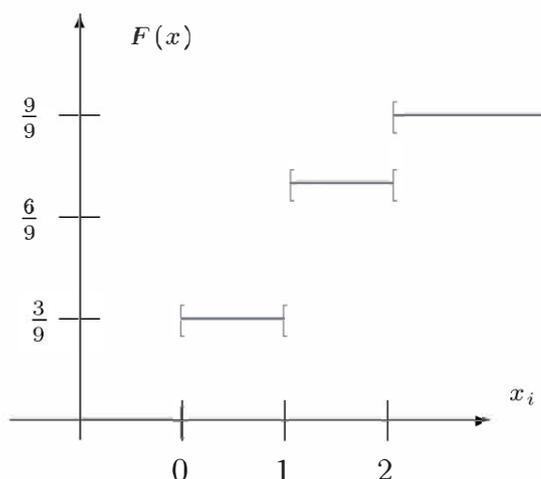


Figure 5-3 Distribution de E

La **fonction de répartition** d'une variable aléatoire est la fonction :

$$F(x) = P(X \leq x)$$

Cette définition indique que les probabilités sont cumulées dans la fonction de répartition ; ainsi cette fonction est croissante et comprise entre 0 et 1. Dans le cas d'une variable discrète, il s'agit d'une fonction en escalier.

Figure 5-4 Fonction de répartition de E

b) Variable continue

Loi et représentation

Les variables continues sont des variables pour lesquelles toutes les valeurs sont possibles dans un intervalle donné.

Exemple. Pour le temps d'attente à une caisse, on peut raisonnablement admettre que ce temps est compris entre 0 et une heure et que tous les temps d'attente sont envisageables entre 0 et 1 heure.

Conséquence : la variable temps d'attente peut prendre une infinité de valeurs car on peut diviser le temps en quantités aussi petites que l'on veut. Sachant qu'il y a une infinité de valeurs, chacune d'entre elles est de probabilité nulle. La probabilité d'attendre précisément deux minutes comme la probabilité d'attendre 2,00000001 minutes est nulle. On s'intéresse donc aux probabilités d'être dans un intervalle, par exemple la probabilité d'attendre entre 2 et 4 minutes.

Remarque : une probabilité nulle ne signifie pas que l'événement est impossible. On peut attendre deux minutes mais la probabilité d'attendre deux minutes est nulle.

Le graphique suivant représente la répartition des probabilités d'une variable continue en fonction des valeurs de x_i . Considérons un premier découpage donné par les rectangles rose, par analogie avec les histogrammes la surface de ces rectangles représente la probabilité d'être dans un intervalle donné. On divise la largeur de ces intervalles en deux et on obtient le diagramme rouge qui donne une information plus précise sur la répartition des probabilités de cette variable.

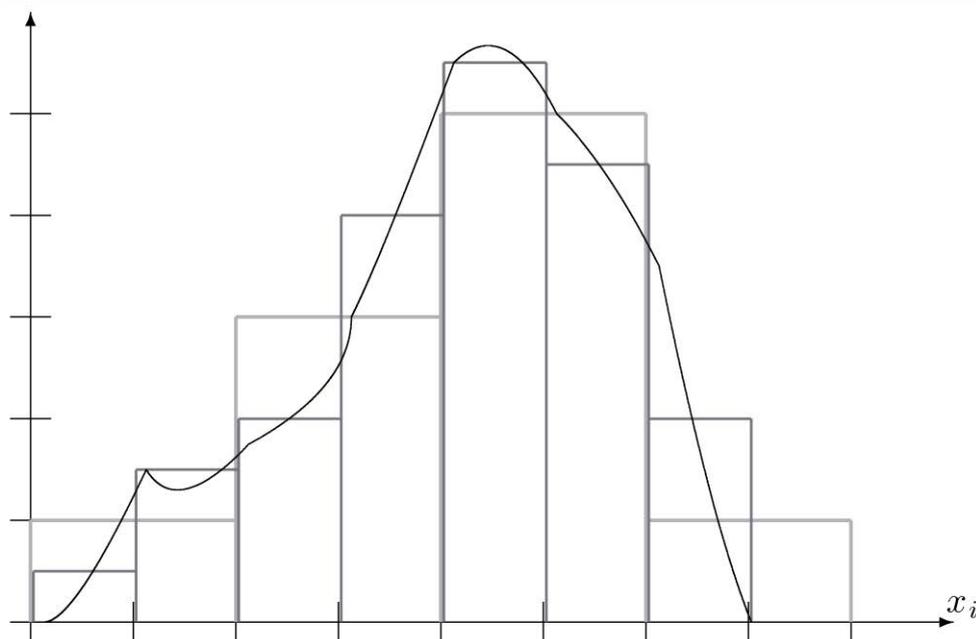


Figure 5-5 Construction de la densité

On pourrait encore diviser les intervalles en une partition fine, au point de ne plus avoir une succession de rectangles mais une courbe continue. Cette courbe (en noir sur la figure) est la **fonction de densité** de la variable aléatoire.

La fonction de densité est une fonction f qui permet de définir la loi d'une variable aléatoire X . Elle vérifie deux critères :

- $f(x) \geq 0$ pour tout x dans Ω .
- $\int_{\Omega} f(x)dx = 1$ autrement dit la somme des probabilités est égale à 1.

Elle permet de définir la **fonction de répartition** :

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t)dt$$

Remarque : la fonction de répartition est positive et croissante.

On a les propriétés suivantes :

- $P(X > x) = \int_x^{+\infty} f(t)dt = 1 - F(x)$
- $P(a \leq X \leq b) = \int_a^b f(t)dt = F(b) - F(a)$

Espérance et écart-type

La densité permet aussi de calculer l'espérance et la variance d'une variable aléatoire.

$$\text{Formule de l'espérance : } E(X) = \int_{-\infty}^{+\infty} xf(x)dx$$

La **variance** se calcule par la formule : $V(X) = E(X^2) - (E(X))^2$ avec

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)dx.$$

On trouve l'écart-type par la formule : $\sigma(X) = \sqrt{V(X)}$.

c) Propriétés de l'espérance et de la variance

Ces propriétés sont valables pour des variables continues et pour des variables discrètes. On note X et Y deux variables aléatoires et a et b deux nombres réels. On a les propriétés suivantes :

1. Espérance

$$E(X + Y) = E(X) + E(Y)$$

$$E(aX + b) = aE(X) + b$$

$$E(aX + bY) = aE(X) + bE(Y)$$

2. Variance

$$V(X + Y) = V(X) + V(Y)$$

$$V(aX + b) = a^2V(X)$$

$$V(aX + bY) = a^2V(X) + b^2V(Y)$$

3. Écart-type

$$\sigma(X + Y) = \sqrt{\sigma^2(X) + \sigma^2(Y)}$$

$$\sigma(aX + b) = |a|\sigma(X)$$

$$\sigma(aX + bY) = \sqrt{a^2\sigma^2(X) + b^2\sigma^2(Y)}$$

Les propriétés de l'espérance sont linéaires. Si on ajoute un nombre à toutes les valeurs de la variable alors l'espérance augmente de ce nombre. Si on multiplie toutes les valeurs de la variable par un nombre alors l'espérance est multipliée par ce nombre. Ceci semble naturel car l'espérance est une mesure de tendance centrale et doit donc être « au centre » des valeurs de la variable.

La variance est un intermédiaire de calcul qui permet d'obtenir l'écart-type. L'écart-type mesure la dispersion d'une variable, par conséquent

les propriétés de la variance et de l'écart-type ne sont pas linéaires. En effet, si on ajoute un nombre à toutes les valeurs de la variable, il y a une translation des valeurs mais aucune modification de la dispersion, donc aucune modification de la variance. Par contre, si on multiplie les valeurs de la variable par un nombre positif, on a une augmentation de la dispersion si ce nombre est plus grand que 1 ou une réduction si ce nombre est plus petit que 1.

Remarque : pour la somme de 2 variables aléatoires, la propriété $V(X + Y) = V(X) + V(Y)$ n'implique pas $\sigma(X + Y) = \sigma(X) + \sigma(Y)$, car mathématiquement $\sqrt{a + b} \neq \sqrt{a} + \sqrt{b}$. Les propriétés de l'écart-type se déduisent de celles de la variance par passage à la racine carrée.



POINTS CLEFS

► Modes de dénombrement

Mode	Contexte d'utilisation	Formule
p-liste	Tirages successifs avec remise	n^p
Arrangements	Tirages successifs sans remise	$\frac{n!}{(n-p)!} = A_n^p$
Combinaisons	Tirages simultanés	$\binom{n}{p} = \frac{n!}{p!(n-p)!}$

► Propriétés des probabilités : on considère deux événements A et B .

$$P(\Omega) = 1 \text{ et } P(\emptyset) = 0$$

$$P(\bar{A}) = 1 - P(A)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$\overline{A \cup B} = \bar{A} \cap \bar{B} \text{ et } \overline{A \cap B} = \bar{A} \cup \bar{B}$$

► Equiprobabilité : dans le cas où chaque événement élémentaire a la même probabilité d'apparition, on calcule les probabilités par la formule :

$$\frac{\text{Nombre de cas favorables}}{\text{Nombre de cas possibles}}$$

- **Probabilités conditionnelles** : la probabilité de A sachant B est donnée par $P_B(A) = \frac{P(A \cap B)}{P(B)}$. Pour éviter une formalisation trop poussée on peut utiliser la représentation sous forme d'**arbre de probabilités**. Dans le cas d'événements indépendants, on a $P(A \cap B) = P(A) \times P(B)$.

► **Lois de probabilités**

- Dans le cas discret : la loi de probabilité d'une variable aléatoire discrète est l'association des valeurs possibles de la variable et de leurs probabilités associées. On peut donner la loi de probabilité sous forme de tableau lorsque l'ensemble des possibles est fini ; sinon on explicite la formule $P(X = x_i)$.

- Dans le cas continu : la donnée de la densité f suffit à définir la loi de probabilité car elle permet le calcul de

$$P(a \leq X \leq b) = \int_a^b f(x)dx \text{ quels que soient } a \text{ et } b.$$

► **Espérance et écart-type**

- L'**espérance** est un indicateur de tendance centrale. Elle se calcule par la formule $\sum_{i \in \Omega} x_i \times p_i$ pour les variables discrètes et par la for-

mule $\int_{-\infty}^{+\infty} xf(x)dx$ pour les variables continues.

- L'**écart-type** est un indicateur de dispersion. L'écart-type est la racine carrée de la variance. La variance est donnée par la relation $V(X) = E(X^2) - (E(X))^2$ où $E(X^2)$ se calcule par la formule $\sum_{i \in \Omega} x_i^2 \times p_i$ pour les variables discrètes et par la formule

$$\int_{-\infty}^{+\infty} x^2 f(x)dx \text{ pour les variables continues.}$$

– **Propriétés**

$$E(aX + b) = aE(X) + b$$

$$E(X + Y) = E(X) + E(Y)$$

$$V(aX + b) = a^2V(X)$$

$$V(X + Y) = V(X) + V(Y)$$

EXERCICES

5.1 Centres de recherche

Une entreprise a ses sites de productions répartis sur trois pays : A, B et C. Cette entreprise a 4 sites de production dans le pays A, 6 dans le pays B et 9 dans le pays C. On cherche à sélectionner trois sites pour y installer des centres de recherche.

- De combien de façon peut-on choisir ces trois sites ?
- Quelle est la probabilité que chaque pays dispose d'un centre de recherche ?
- Quelle est la probabilité qu'un seul pays dispose de tous les centres de recherche ?
- Quelle est la probabilité que seulement deux pays disposent de centres de recherche ?

5.2 Gestion de ressources humaines

Huit personnes postulent pour un poste, le responsable des ressources humaines doit les faire passer successivement en entretien.

- De combien de façons peut-il choisir de faire passer les différents postulants ?
- Le responsable choisit préalablement de ne sélectionner que cinq postulants parmi les huit pour leur faire passer ensuite l'entretien. De combien de façons ce responsable peut-il faire ces cinq entretiens ?

5.3 Cours boursiers

On remarque que les conseils des analystes financiers ont une influence sur l'évolution des cours. Pour un titre donné, lorsque le conseil est à l'achat, le cours monte dans 60 % des cas. Lorsque le cours monte, on constate qu'un conseil d'achat a été donné dans 70 % des cas. On note que pour 75 % des titres, le conseil est à l'achat.

- Quelle est la probabilité que le cours d'un titre monte ?
- Quelle est la probabilité qu'un cours baisse ou stagne si on a un conseil d'achat ?

5.4 Influence du soleil

Le soleil est supposé avoir une influence sur le comportement des traders. Le tableau page suivante montre les résultats constatés sur plusieurs jours d'observations.

	Progression des cours	Régression ou stagnation des cours
Journée ensoleillée	48 %	12 %
Journée sans soleil	32 %	8 %

- a) Quelle est la probabilité de voir progresser les cours ?
- b) Quelle est la probabilité de voir progresser les cours sachant que la journée est ensoleillée ?
- c) Que peut-on conclure ?

5.5 Location de véhicules

Une société loue des véhicules de transport. Elle dispose de 6 véhicules qui passent régulièrement en révision. On note X la variable aléatoire associée au nombre de véhicules disponibles un jour donné. On observe les résultats suivants :

x_i	0	1	2	3	4	5
p_i	5 %	6 %	14 %	18 %	25 %	32 %

La demande journalière en véhicule varie d'un jour à l'autre. On note Y la variable aléatoire représentant la demande en véhicules.

On observe les résultats suivants :

y_i	0	1	2	3	4	5	6
p_i	3 %	7 %	10 %	24 %	22 %	26 %	8 %

Les variables X et Y sont supposées indépendantes. On note N la variable aléatoire associée au nombre de véhicules effectivement loués.

- a) Déterminer la loi de N .
- b) Donner son espérance et sa variance.
- c) Si un véhicule est loué 500 euros. Déterminer l'espérance et la variance du gain réalisé par les locations.

5.6 Écart de travail

Deux commerciaux effectuent plusieurs visites dans une journée de travail. Le nombre de visites effectuées par chacun suit une loi X donnée

dans le tableau suivant. Les deux commerciaux travaillent de façon indépendante : le nombre de visites de l'un n'a aucun impact sur le nombre de visites de l'autre. On s'intéresse à l'écart Y entre le nombre de visites effectuées par chacun des deux commerciaux.

x_i	5	6	7	8	9	10
p_i	4 %	12 %	20 %	30 %	12 %	22 %

- Donner la loi de l'écart Y .
- Quelle est la probabilité d'avoir un écart de visites strictement supérieur à 3 entre les deux commerciaux ?
- Donner l'espérance et l'écart-type de Y .

5.7 Utilisation des définitions

Une variable aléatoire prend ses valeurs dans l'ensemble $\{-1 ; 0 ; 2\}$, son espérance est 0 et sa variance est 1. Déterminer la loi de cette variable.

5.8 Échantillon représentatif ?

Une usine fabrique des pièces dont 2 % sont défectueuses. On dispose d'une production de 4 000 pièces. On prélève au hasard 3 pièces. On note X le nombre de pièces défectueuses dans l'échantillon.

- Déterminer la loi de X .
- Quelle est la proportion moyenne de pièces défectueuses dans l'échantillon ?
- Un inspecteur teste la production jusqu'à trouver une pièce défectueuse. Donner la loi de Y , le rang d'apparition de la première pièce défectueuse sachant qu'une pièce testée n'est pas remise dans la production.

5.9 Loi continue

Dans les portefeuilles boursiers, la proportion du titre A est donné par une variable aléatoire X de densité $f(x) = 2x$.

- Montrer que f est bien une densité en précisant son ensemble de définition.
- Donner la proportion moyenne du titre A dans les portefeuilles.
- Donner l'écart-type de X .
- Représenter graphiquement la densité de X et sa fonction de répartition.

SOLUTIONS

Exercice 5.1

a) Il y a en tout 19 sites de productions sur l'ensemble des trois pays. On doit choisir de sélectionner trois sites parmi ces 19. Il s'agit d'un cas de « tirage simultané ». Le nombre de choix possible est donc :

$$\binom{19}{3} = \frac{19!}{3!(19-3)!} = \frac{19 \times 18 \times 17}{3 \times 2} = 969$$

b) Pour que chaque pays dispose d'un centre de recherche, le choix d'un site se fait séparément dans chaque pays. Dans le pays A, il y a 4 choix possibles ; dans le pays B, 6 choix possibles ; et dans le pays C, 9 choix possibles. Au total, il y a $4 \times 6 \times 9 = 216$ façons de choisir les trois centres. Comme il s'agit d'un cas d'équiprobabilité la formule $\frac{\text{Nombre de cas favorables}}{\text{Nombre de cas possibles}}$ indique une probabilité

$$P_1 = \frac{216}{969} \approx 22,29 \%$$

c) Pour qu'un seul pays dispose de tous les centres de recherches, on dispose soit tous les centres dans le pays A, soit tous les centres dans le pays B, soit tous les centres dans le pays C.

Dans le pays A, il y a $\binom{4}{3} = 4$ façons de choisir les trois centres.

$$\text{Dans le pays B, } \binom{6}{3} = \frac{6!}{3!(6-3)!} = \frac{6 \times 5 \times 4}{3 \times 2} = 20$$

$$\text{Dans le pays C, } \binom{9}{3} = \frac{9!}{3!(9-3)!} = \frac{9 \times 8 \times 7}{3 \times 2} = 84$$

Il y a donc en tout $4 + 20 + 84 = 108$ façons de choisir les trois centres. La probabilité de trouver tous les centres dans le même pays est donc

$$P_2 = \frac{108}{969} \approx 11,15 \%$$

d) Cette probabilité est difficile à déterminer directement, cependant on peut constater que : soit tous les centres sont dans un seul pays, soit tous les centres sont dans deux pays différents, soit les centres sont dans trois pays différents. Il n'y a pas d'autres alternatives. Ainsi, si on note P_3 la probabilité que les centres soient dans deux pays différents, on a

$$P_1 + P_2 + P_3 = 1, \text{ d'où } P_3 = 1 - \frac{216}{969} - \frac{108}{969} = \frac{215}{323} \approx 66,56 \%$$

Exercice 5.2

a) Les huit postulants passent successivement. Pour le premier entretien il y a donc huit choix possibles, pour le second il n'y en a plus que 7, pour le troisième plus que 6 et ainsi de suite. On trouve donc en tout $8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 8! = 40\,320$ façons de faire passer les huit entretiens.

b) Le responsable sélectionne dans un premier temps cinq postulants parmi les huit. Il y a $\binom{8}{5}$ façons de le faire. Ensuite, il fait passer successivement les entretiens aux cinq postulants ; il y a donc $5!$ façons de le faire.

$$\begin{aligned} \text{Il y a ainsi en tout } \binom{8}{5} \times 5! &= \frac{8!}{5! \times (8-5)!} \times 5! = \frac{8!}{(8-5)!} \\ &= 8 \times 7 \times 6 \times 5 \times 4 = 6\,720 \end{aligned}$$

Soit 6 720 façons de faire passer ces cinq entretiens. On remarque qu'il s'agit aussi du nombre d'arrangements à 5 éléments dans un ensemble à 8 éléments.

Exercice 5.3

a) Modélisons l'énoncé par des événements et des probabilités. Notons A l'événement « le conseil est à l'achat » et M « le cours monte ». On a alors $P_A(M) = 0,6$; $P_M(A) = 0,7$ et $P(A) = 0,75$.

D'après la formule des probabilités conditionnelles, on a :

$$P_A(M) = \frac{P(A \cap M)}{P(A)}$$

$$\text{Ainsi } 0,6 = \frac{P(A \cap M)}{0,75}, \text{ d'où } P(A \cap M) = 0,6 \times 0,75$$

$$\text{De plus } P_M(A) = \frac{P(A \cap M)}{P(M)}, \text{ donc } 0,7 = \frac{0,6 \times 0,75}{P(M)}$$

$$\text{Ainsi } P(M) = \frac{0,6 \times 0,75}{0,7} \approx 64,29 \%$$

b) On cherche la probabilité $P_A(\bar{M})$. La formule des probabilités conditionnelles indique que $P_A(\bar{M}) = \frac{P(A \cap \bar{M})}{P(A)}$. Il reste à trouver

$P(A \cap \bar{M})$. Les événements M et \bar{M} sont complémentaires, ainsi $P(A) = P(A \cap M) + P(A \cap \bar{M})$.

$$\text{Donc } P(A \cap \bar{M}) = 0,75 - 0,75 \times 0,6$$

$$\text{On a donc } P_A(\bar{M}) = \frac{0,75 - 0,75 \times 0,6}{0,75} = 0,4$$

Exercice 5.4

a) Notons C l'événement « une progression des cours » et S l'événement « une journée ensoleillée ». On a :

$$P(C) = P(C \cap S) + P(C \cap \bar{S}) = 48 \% + 32 \% = 80 \%$$

b) On cherche $P_S(C) = \frac{P(S \cap C)}{P(S)} = \frac{48 \%}{48 \% + 12 \%} = 80 \%$

c) On remarque que $P(C) = P_S(C)$ ainsi les événements S et C sont indépendants.

Exercice 5.5

a) Dans un premier temps, déterminons l'ensemble des valeurs prises par N . Dans le tableau suivant, la première ligne représente les valeurs possibles de X et la première colonne les valeurs possibles de Y . L'intérieur du tableau est l'ensemble des valeurs de N .

	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	1	1	1	1
2	0	1	2	2	2	2
3	0	1	2	3	3	3
4	0	1	2	3	4	4
5	0	1	2	3	4	5
6	0	1	2	3	4	5

Par exemple, si $x_i = 3$ et $y_i = 5$, l'offre est de 3 et la demande est de 5. La société ne peut louer que 3 véhicules, donc $n_i = 3$. On remarque de manière générale que $N = \text{Min}(X, Y)$.

- $P(N = 5) = P(X = 5 \cap Y = 5) + P(X = 5 \cap Y = 6)$ d'après le tableau. Comme X et Y sont indépendantes, on a :

$$\begin{aligned} P(N = 5) &= P(X = 5) \times P(Y = 5) + P(X = 5) \times P(Y = 6) \\ &= 0,32 \times 0,26 + 0,32 \times 0,08 = 10,88 \% \end{aligned}$$

- De la même manière :

$$\begin{aligned} P(N = 4) &= P(X = 4) \times P(Y = 4) + P(X = 4) \times P(Y = 5) \\ &\quad + P(X = 4) \times P(Y = 6) + P(X = 5) \times P(Y = 4) \\ &= 0,25 \times 0,22 + 0,25 \times 0,26 + 0,25 \times 0,08 + 0,32 \times 0,22 \\ &= 21,04 \% \end{aligned}$$

- $P(N = 3) = 0,18 \times 0,24 + 0,18 \times 0,22 + 0,18 \times 0,26 + 0,18 \times 0,08 + 0,25 \times 0,24 + 0,32 \times 0,24 = 28,08 \%$
- $P(N = 2) = 0,14 \times 0,1 + 0,14 \times 0,24 + 0,14 \times 0,22 + 0,14 \times 0,26 + 0,14 \times 0,08 + 0,18 \times 0,1 + 0,25 \times 0,1 + 0,32 \times 0,1 = 20,1 \%$
- $P(N = 1) = 0,06 \times 0,07 + 0,06 \times 0,1 + 0,06 \times 0,24 + 0,06 \times 0,22 + 0,06 \times 0,26 + 0,06 \times 0,08 + 0,14 \times 0,07 + 0,18 \times 0,07 + 0,25 \times 0,07 + 0,32 \times 0,07 = 12,05 \%$
- $P(N = 0) = 0,05 + 0,03 \times (0,06 + 0,14 + 0,18 + 0,25 + 0,32) = 7,85 \%$

On résume la loi de N ainsi :

x_i	0	1	2	3	4	5
p_i	7,85 %	12,05 %	20,1 %	28,08 %	21,04 %	10,88 %

On remarque que la somme des probabilités est égale à 1.

b) L'espérance :

$$E(N) = 0 \times 0,0785 + 1 \times 0,1205 + 2 \times 0,201 + 3 \times 0,2808 + 4 \times 0,2104 + 5 \times 0,1088 = 2,7505$$

La variance : $V(N) = E(N^2) - E(N)^2$

On a :

$$E(N^2) = 0^2 \times 0,0785 + 1^2 \times 0,1205 + 2^2 \times 0,201 + 3^2 \times 0,2808 + 4^2 \times 0,2104 + 5^2 \times 0,1088 = 9,5381$$

$$\text{et } E(N)^2 = 2,7505^2 = 7,56525$$

$$\text{ainsi } V(N) = 9,5381 - 7,56525 = 1,97285$$

c) Le gain est proportionnel au nombre de véhicules loués. On a $G = 500 \times N$. Ainsi :

$$E(G) = 500 \times E(N) = 500 \times 2,7505 = 1\,375,25$$

$$\text{et } V(G) = 500^2 \times V(N) = 500^2 \times 1,97285 = 493\,212,44$$

Exercice 5.6

a) Voyons dans un premier temps les valeurs possibles de l'écart Y entre les deux commerciaux. Dans le tableau suivant, la première ligne est donnée par le nombre de visites du premier commercial et la première colonne est donnée par le nombre de visites du second commercial. L'intérieur du tableau est constitué des valeurs de l'écart Y .

	5	6	7	8	9	10
5	0	1	2	3	4	5
6	1	0	1	2	3	4
7	2	1	0	1	2	3
8	3	2	1	0	1	2
9	4	3	2	1	0	1
10	5	4	3	2	1	0

Par exemple, si le premier commercial fait 4 visites et le second commercial fait 2 visites, l'écart est de 2.

Notons X_1 la variable représentant le premier commercial et X_2 la variable représentant le second commercial. Comme ils suivent chacun la même loi, on a $X = X_1 = X_2$.

Ainsi :

$$\begin{aligned} \bullet P(Y = 5) &= P(X_1 = 5 \cap X_2 = 10) + P(X_2 = 5 \cap X_1 = 10) \\ &= 2 \times P(X_1 = 5 \cap X_2 = 10) \end{aligned}$$

Comme ils suivent la même loi et comme X_1 et X_2 sont indépendantes :

$$P(Y = 5) = 2 \times P(X = 5) \times P(X = 10) = 2 \times 0,04 \times 0,22 = 1,76 \%$$

$$\begin{aligned} \bullet P(Y = 4) &= 2 \times (P(X = 5) \times P(X = 9) + P(X = 6) \times P(X = 10)) \\ &= 2 \times (0,04 \times 0,12 + 0,12 \times 0,22) = 6,24 \% \end{aligned}$$

$$\begin{aligned} \bullet P(Y = 3) &= 2 \times (P(X = 5) \times P(X = 8) + P(X = 6) \times P(X = 9) \\ &\quad + P(X = 7) \times P(X = 10)) \\ &= 2 \times (0,04 \times 0,3 + 0,12 \times 0,12 + 0,2 \times 0,22) \\ &= 14,08 \% \end{aligned}$$

$$\begin{aligned} \bullet P(Y = 2) &= 2 \times (P(X = 5) \times P(X = 7) + P(X = 6) \times P(X = 8) \\ &\quad + P(X = 7) \times P(X = 9) + P(X = 8) \times P(X = 10)) \\ &= 2 \times (0,04 \times 0,2 + 0,12 \times 0,3 + 0,2 \times 0,12 + 0,3 \times 0,22) \\ &= 26,8 \% \end{aligned}$$

$$\begin{aligned} \bullet P(Y = 1) &= 2 \times (P(X = 5) \times P(X = 6) + P(X = 6) \times P(X = 7) \\ &\quad + P(X = 7) \times P(X = 8) + P(X = 8) \times P(X = 9) \\ &\quad + P(X = 9) \times P(X = 10)) \\ &= 2 \times (0,04 \times 0,12 + 0,12 \times 0,2 + 0,2 \times 0,3 \\ &\quad + 0,3 \times 0,12 + 0,12 \times 0,22) \\ &= 30,24 \% \end{aligned}$$

$$\begin{aligned}
 \bullet P(Y = 0) &= P(X = 5)^2 + P(X = 6)^2 + P(X = 7)^2 + P(X = 8)^2 \\
 &\quad + P(X = 9)^2 + P(X = 10)^2 \\
 &= 0,04^2 + 0,12^2 + 0,2^2 + 0,3^2 + 0,12^2 + 0,22^2 \\
 &= 20,88 \%
 \end{aligned}$$

y_i	0	1	2	3	4	5
p_i	20,88 %	30,24 %	26,8 %	14,08 %	6,24 %	1,76 %

On résume la loi de Y ainsi :

$$\begin{aligned}
 \text{b) On cherche } P(Y > 3) &= P(Y = 4) + P(Y = 5) \\
 &= 6,24 \% + 1,76 \% = 8 \%
 \end{aligned}$$

$$\begin{aligned}
 \text{c) Espérance : } E(Y) &= 0 \times 0,2088 + 1 \times 0,3024 + 2 \times 0,268 \\
 &\quad + 3 \times 0,1408 + 4 \times 0,0624 + 5 \times 0,0176 = 1,5984
 \end{aligned}$$

$$\text{Écart-type : } \sigma(Y) = \sqrt{E(Y^2) - E(Y)^2}$$

$$\begin{aligned}
 E(Y^2) &= 0^2 \times 0,2088 + 1^2 \times 0,3024 + 2^2 \times 0,268 + 3^2 \times 0,1408 \\
 &\quad + 4^2 \times 0,0624 + 5^2 \times 0,0176 = 4,08
 \end{aligned}$$

$$\text{et } E(Y)^2 = 1,5984^2 = 2,55488$$

$$\text{d'où } \sigma(Y) = \sqrt{4,08 - 2,55488} = 1,23496$$

Exercice 5.7

On cherche à donner la loi de cette variable (notée X). Pour cela posons $x = P(X = -1)$, $y = P(X = 0)$ et $z = P(X = 2)$.

On a $E(X) = 0$, donc $-1 \times x + 0 \times y + 2 \times z = 0$, soit $-x + 2z = 0$

On a $V(X) = 1$, donc $E(X^2) - E(X)^2 = 1$

Or $E(X) = 0$, d'où $E(X^2) = 1$.

Ainsi $(-1)^2 \times x + 0^2 \times y + 2^2 \times z = 1$, soit $x + 4z = 1$

En additionnant les deux lignes précédentes, on trouve $6z = 1$, d'où

$$z = \frac{1}{6}$$

On en déduit que comme $x = 2z$, on a $x = \frac{1}{3}$

Comment trouver y ? Il ne faut pas oublier que la somme des probabilités fait 1, donc $x + y + z = 1$, soit $\frac{1}{3} + y + \frac{1}{6} = 1$ et par conséquent

$$y = \frac{1}{2}$$

La loi de X se résume ainsi :

x_i	- 1	0	2
p_i	1/3	1/2	1/6

Exercice 5.8

a) Il y a dans la production en tout $2\% \times 4\,000 = 80$ pièces défectueuses. Le nombre de façons de choisir au hasard 3 pièces dans l'ensemble de la production est :

$$\binom{4\,000}{3} = \frac{4\,000!}{3! \times (4\,000 - 3)!} = \frac{4\,000 \times 3\,999 \times 3\,998}{2 \times 3} = 10\,658\,668\,000$$

Car il s'agit d'un tirage simultané.

Dans l'échantillon de 3 pièces, il peut y avoir 0, 1, 2 ou 3 pièces défectueuses. Voyons les différents cas :

- Pour $P(X = 0)$, il n'y a aucune pièce défectueuse dans l'échantillon donc l'échantillon a été prélevé parmi les 3 920 pièces non défectueuses. Le nombre de façon de prélever 3 pièces non défectueuses est :

$$\begin{aligned} \binom{3\,920}{3} &= \frac{3\,920!}{3! \times (3\,920 - 3)!} = \frac{3\,920 \times 3\,919 \times 3\,918}{2 \times 3} \\ &= 10\,031\,699\,440 \end{aligned}$$

$$\text{Ainsi } P(X = 0) = \frac{10\,031\,699\,440}{10\,658\,668\,000} \approx 94,12\%$$

- Pour $P(X = 1)$, on prélève deux pièces parmi les non défectueuses, soit $\binom{3\,920}{2} = \frac{3\,920!}{2! \times (3\,920 - 2)!} = \frac{3\,920 \times 3\,919}{2} = 7\,681\,240$ choix possibles et une pièce parmi les défectueuses, soit 80 choix possibles. Ainsi $P(X = 1) = \frac{7\,681\,240 \times 80}{10\,658\,668\,000} \approx 5,77\%$

- Pour $P(X = 2)$, on prélève une pièces parmi les non défectueuses, soit 3 920 choix possibles, et deux pièces parmi les défectueuses, soit

$$\binom{80}{2} = \frac{80!}{2! \times (80 - 2)!} = \frac{80 \times 79}{2} = 3\,160 \text{ choix possibles.}$$

$$\text{Ainsi } P(X = 2) = \frac{3\,920 \times 3\,160}{10\,658\,668\,000} \approx 0,12\%$$

- Pour $P(X = 3)$ on prélève 3 pièces parmi les défectueuses, soit

$$\binom{80}{3} = \frac{80!}{3! \times (80 - 3)!} = \frac{80 \times 79 \times 78}{2 \times 3} = 82\,160.$$

Ainsi $P(X = 3) = \frac{82\,160}{10\,658\,668\,000} \approx 7,7 \times 10^{-6}$ ce qui est négligeable devant les autres probabilités.

On résume donc la loi X de façon approchée ainsi :

x_i	- 1	0	2	3
p_i	94,12 %	5,77 %	0,12 %	0 %

b) Calculons dans un premier temps le nombre moyen de pièces défectueuses dans l'échantillon. Il s'agit de :

$$E(X) = 0 \times 0,9412 + 1 \times 0,0577 + 2 \times 0,0012 + 3 \times 0 = 0,0601$$

Sachant que l'échantillon contient trois éléments la proportion moyenne de pièces défectueuses est $\frac{0,0601}{3} \approx 2\%$. On retrouve le pourcentage de l'ensemble de la production.

c) Les valeurs possibles pour Y sont $\{1, 2, 3, \dots, 3\,921\}$. En effet, il y a en tout 4 000 pièces dont 80 défectueuses, et on peut en trouver une dès le premier tirage. Au plus tard lors du tirage 3 921, on va trouver une pièce défectueuse si toutes les précédentes ne le sont pas.

On a :

- $P(Y = 1)$ = Probabilité de tirer une pièce défectueuse dès le premier tirage $= \frac{80}{4\,000}$

- $P(Y = 2)$ = Probabilité de tirer une pièce non défectueuse au premier tirage et une pièce défectueuse au second tirage $= \frac{3\,920}{4\,000} \times \frac{80}{3\,999}$

- $P(Y = 3) = \frac{3\,920 \times 3\,919 \times 80}{4\,000 \times 3\,999 \times 3\,998}$

- $P(Y = 4) = \frac{3\,920 \times 3\,919 \times 3\,918 \times 80}{4\,000 \times 3\,999 \times 3\,998 \times 3\,997}$

- Il serait trop long de donner l'ensemble des probabilités associées aux valeurs de la variable Y , on identifie donc les calculs précédents à une formule faisant intervenir les arrangements :

$$P(Y = k) = \frac{A_{3\,920}^{k-1} \times 80}{A_{4\,000}^k}$$

Exercice 5.9

a) Comme x est une proportion d'un titre dans un portefeuille, on a $0 \leq x \leq 1$. Ainsi l'ensemble de définition de f est $[0;1]$. Pour montrer que f est une densité, il faut montrer que cette fonction est positive sur son ensemble de définition et que son intégrale vaut 1.

• On a $0 \leq x$, ainsi $0 \leq 2x$ et $0 \leq f(x)$

$$\text{et } \int_0^1 f(x)dx = \int_0^1 2xdx = [x^2]_0^1 = 1^2 - 0^2 = 1$$

f est bien une densité.

b) Il s'agit du calcul de l'espérance de la variable. On a :

$$E(X) = \int_0^1 xf(x)dx = \int_0^1 2x^2dx = \left[2\frac{x^3}{3}\right]_0^1 = 2\frac{1^3}{3} - 2\frac{0^3}{3} = \frac{2}{3}$$

c) L'écart-type de X se calcule par la formule :

$$\sigma(X) = \sqrt{E(X^2) - E(X)^2}$$

$$\begin{aligned} \text{Or } E(X^2) &= \int_0^1 x^2 f(x)dx = \int_0^1 2x^3dx \\ &= \left[2\frac{x^4}{4}\right]_0^1 = \frac{1^4}{2} - \frac{0^4}{2} = \frac{1}{2} \end{aligned}$$

$$\text{On a donc } \sigma(X) = \sqrt{\frac{1}{2} - \left(\frac{2}{3}\right)^2} \approx 0,2357.$$

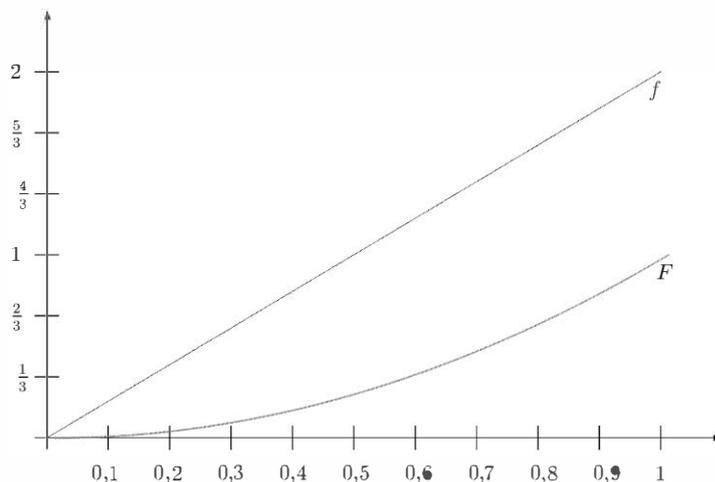
d) Notons $F(x)$ la fonction de répartition de X . On a par définition

$$F(x) = \int_0^x f(t)dt = \int_0^x 2tdt = [t^2]_0^x = x^2 \text{ pour } x \text{ compris entre } 0 \text{ et } 1.$$

Pour $x < 0$, on a $F(x) = 0$ et pour $x > 1$, on a $F(x) = 1$.

La fonction f est une fonction linéaire sur $[0;1]$.

Voici la représentation graphique de ces deux fonctions sur $[0,1]$:



Lois fondamentales de probabilités

OBJECTIFS

- ▶ Connaître la définition des lois binomiale, de Poisson et normale
- ▶ Savoir calculer des probabilités avec ces lois
- ▶ Connaître les représentations graphiques des distributions de ces lois
- ▶ Savoir utiliser les tables de la loi Normale centrée réduite
- ▶ Connaître les approximations de lois
- ▶ Savoir modéliser une expérience par l'une de ces trois lois

PLAN

- 6.1 Loi binomiale
- 6.2 Loi de Poisson
- 6.3 Loi normale
- 6.4 Approximations de lois

Les expériences observées en économie ou en finance nécessitent une modélisation par des lois de probabilités. À chaque expérience pourrait correspondre une loi, cependant certaines familles d'expériences conduisent à des modélisations similaires. On limitera donc l'étude à certaines lois « fondamentales », autrement dit à des lois très souvent utilisées dans les modèles. Ces lois sont la loi binomiale dans le cas fini discret, la loi de Poisson dans le cas infini discret et la loi normale dans le cas continu. La loi normale est de loin la plus importante car la majorité des lois de probabilités peuvent être approximées par elle.

6.1 LOI BINOMIALE

a) Construction de la loi binomiale

On lance une pièce dont la probabilité d'obtenir pile est p . On note X la variable aléatoire associée au nombre de fois que l'on obtient pile dans cette expérience. Comme on ne lance la pièce qu'une fois, on a seulement deux possibilités : soit on obtient pile et $X = 1$, soit on obtient face et $X = 0$. La loi de X se résume ainsi :

- $P(X = 1) = p$
- $P(X = 0) = 1 - p$

Il s'agit d'un **schéma de Bernoulli**.

$$\text{On a } E(X) = 1 \times p + 0 \times (1 - p) = p$$

$$\text{et } V(X) = 1^2 \times p + 0^2 \times (1 - p) - p^2 = p - p^2 = p(1 - p)$$

On lance maintenant cette même pièce n fois de suite et on s'intéresse toujours à X , le nombre d'obtention de pile. L'ensemble des possibles pour X est $\{0, 1, \dots, n\}$. En effet, il est possible de ne jamais obtenir de pile et $X = 0$, il est possible de n'en obtenir qu'un et $X = 1$, et il est possible d'en obtenir n et $X = n$.

Comment trouver $P(X = k)$ pour $k \in \{0, 1, \dots, n\}$?

L'événement $X = k$ est « l'obtention de k piles parmi n lancers ». Cet événement peut s'obtenir de plusieurs façons :

- $PPP \dots PFFFF \dots F$: k piles suivis de $n - k$ faces.
- $FFF \dots FPP \dots P$: $n - k$ faces suivis de k piles.
- $FPPFP \dots PFF \dots F$: face suivi de pile, suivi de face, suivi de $k - 1$ piles, suivis de $n - k - 2$ faces.

Il y a bien d'autres façons de réaliser l'événement $X = k$. Cherchons à les dénombrer.

On remarque qu'une suite de lancers réalisant $X = k$ est la disposition de k piles dans n emplacements différents (chaque emplacement correspondant à un lancer). Les autres emplacements sont automatiquement occupés par des faces. Il s'agit donc de choisir k emplacements parmi n , c'est le nombre de combinaisons à k éléments dans un ensemble à n

éléments, à savoir :
$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Chaque lancer étant indépendant du précédent, la probabilité d'obtenir un ensemble de k piles et $n - k$ faces est donc : $p^k(1 - p)^{n-k}$

On a donc : $P(X = k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$

Cet exemple permet la construction de la loi **binomiale**.

b) Définition de la loi binomiale

Si X suit une loi **binomiale** de paramètres n et p , on a :

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \text{ pour } k \in \{0, 1, 2, \dots, n\}$$

Remarque : la formule du binôme est $(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}$

Calculons $\sum_{k=0}^n P(X = k)$ pour une loi binomiale.

On a $\sum_{k=0}^n P(X = k) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k}$. En identifiant cette expression à la formule du binôme, on trouve :

$$\sum_{k=0}^n P(X = k) = (p + 1 - p)^n = 1^n = 1$$

Ainsi, on vient de montrer que la somme des probabilités d'une loi binomiale est égale à 1.

Espérance : $E(X) = np$

Variance et écart-type : $V(X) = np(1-p)$ et $\sigma(X) = \sqrt{np(1-p)}$.

Remarque : pour l'espérance et la variance, il suffit de multiplier l'espérance et la variance du schéma de Bernoulli par n .



Dans quels cas utiliser cette loi ?

On utilise cette loi si on cherche à dénombrer les réussites à une suite d'expériences identiques et indépendantes.

Exemples.

- Lancer de dé : nombre d'obtention de 6 en 30 lancers.
- Test d'une production : nombre d'obtention de pièces défectueuses en 50 tests dans une production très grande qui contient 2 % de pièces défectueuses.

– Réponse à un questionnaire : nombre de réponses justes si on répond totalement au hasard à un questionnaire composé de 10 questions pour lesquelles il y a 3 réponses possibles.

c) Représentation graphique

La loi binomiale dépend de deux paramètres, n et p . Voyons graphiquement l'influence de chacun de ces deux paramètres.

Influence de p

Pour observer l'influence de p sur la distribution de la loi binomiale, on fixe $n = 8$ et on fait varier p de 10 % à 50 %.

Le tableau suivant indique les valeurs en pourcentage de $P(X = k)$.

Tableau 6-1 Influence de p

k	0	1	2	3	4	5	6	7	8
$p = 10 \%$	43,05	38,26	14,88	3,31	0,46	0,04	0	0	0
$p = 50 \%$	0,39	3,13	10,94	21,88	27,34	21,88	10,94	3,13	0,39

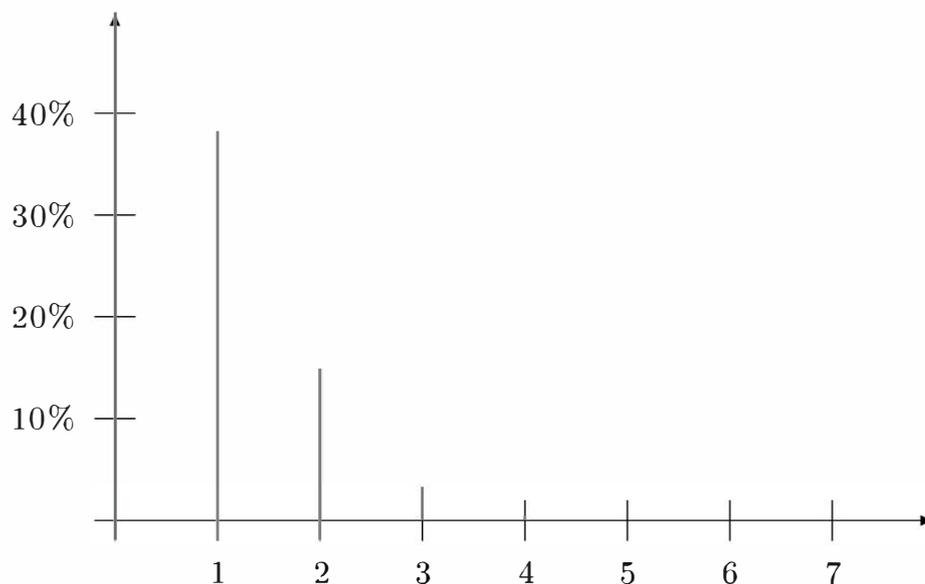


Figure 6-1 Distribution binomiale pour $n = 8$ et $p = 10 \%$

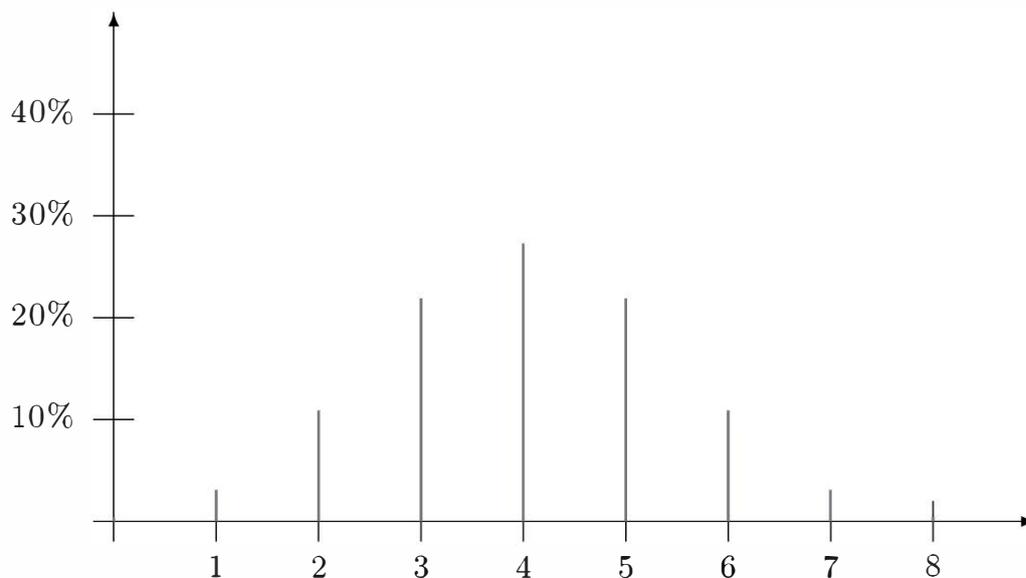


Figure 6-2 Distribution binomiale pour $n=8$ et $p=50\%$

Comparaison des observations : lorsque p est petit, on constate une décroissance de $P(X = k)$, en effet la probabilité de succès est plus faible que celle d'échec. On en déduit que si p était grand, on aurait une croissance de $P(X = k)$. Pour $p = 50\%$, on constate une symétrie de la distribution, ce qui se comprend car il y a la même chance de succès que d'échec.

Remarque : il n'est pas intéressant de dépasser 50% pour p . En changeant de variable, on peut revenir au cas de $p < 50\%$. En effet si le nombre de succès suit une loi binomiale de paramètre $p = 60\%$, le nombre d'échecs suit une loi binomiale de paramètre $p = 40\%$.

Remarque : une distribution à $p = 10\%$ « ressemble » à une distribution de Poisson, alors qu'une distribution à $p = 50\%$ « ressemble » à une distribution normale. Cette remarque sera utile lors des approximations de lois.

Influence de n

Pour observer l'influence de n , on fixe $p = 30\%$ et on fait varier n . Le tableau suivant indique les valeurs en pourcentage de $P(X = k)$ pour $n = 4$ et $n = 8$.

Tableau 6-2 Influence de n

k	0	1	2	3	4	5	6	7	8
$n = 4$	24,01	41,16	26,46	7,56	0,81	0	0	0	0
$n = 8$	5,57	19,77	29,65	25,41	13,61	4,47	1,00	0,12	0

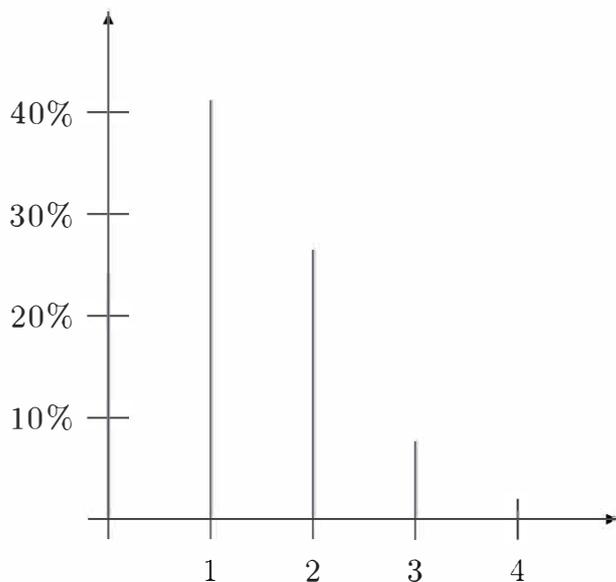


Figure 6-3 Distribution binomiale pour $n = 4$ et $p = 30\%$

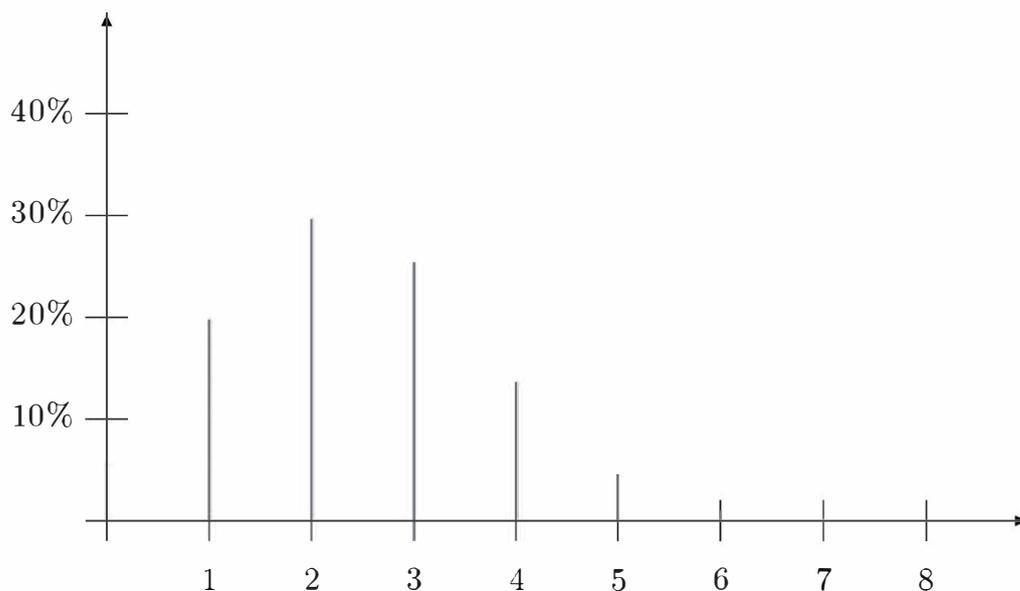


Figure 6-4 Distribution binomiale pour $n = 8$ et $p = 30\%$

Comparaison des observations : lorsque n augmente, la distribution a globalement la même allure mais les résultats sont plus étalés.

Remarque : en étalant les résultats, la distribution ressemble un peu plus à une distribution normale, ce qui sera utile au niveau des approximations de lois.

6.2 LOI DE POISSON

a) Définition de la loi de Poisson

Une variable de **Poisson** prend ses valeurs dans l'ensemble \mathbb{N} . La loi de Poisson se définit à l'aide du seul paramètre λ selon la relation :

$$P(X = k) = e^{-\lambda} \times \frac{\lambda^k}{k!}$$

Cette loi sert de modèle à des expériences. On constate par exemple que les phénomènes d'arrivée de clients dans un magasin ou le nombre de clients à une caisse suivent généralement une loi de Poisson.

Remarque : on ne cherchera donc pas à construire une loi de Poisson à partir des données d'un exercice mais plutôt à l'appliquer quand l'énoncé indique que son utilisation est appropriée, c'est-à-dire si une expérience conduit à constater que la moyenne statistique est proche de la variance.

Espérance et variance : $E(X) = V(X) = \lambda$

Écart-type : $\sigma(X) = \sqrt{\lambda}$

Remarque : l'espérance et la variance d'une loi de Poisson sont égales. Cette propriété est caractéristique d'une loi de Poisson, ainsi dans une expérience, si statistiquement la variance est égale à la moyenne, on peut envisager la modélisation de l'expérience par une loi de Poisson.

b) Représentation graphique

Le tableau et les différents graphes présentent la distribution d'une loi de Poisson pour différentes valeurs de λ : $\lambda = 1$, $\lambda = 2$ et $\lambda = 3,5$.

Tableau 6-3 Influence de λ sur une distribution de Poisson

k	0	1	2	3	4	5	6	7	8	9	10
$\lambda = 1$	36,79	36,79	18,39	6,13	1,53	0,05	0,001	0	0	0	0
$\lambda = 2$	13,53	27,07	27,07	18,05	9,02	3,61	1,20	0,34	0,09	0	0
$\lambda = 3,5$	3,02	10,57	18,50	21,58	18,88	13,22	7,71	3,85	1,68	0,66	0,23

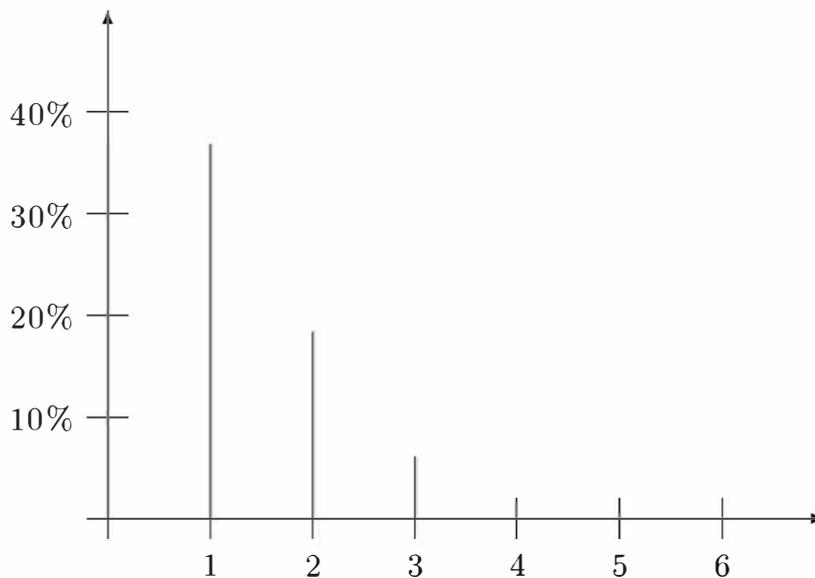


Figure 6-5 Distribution de Poisson pour $\lambda = 1$

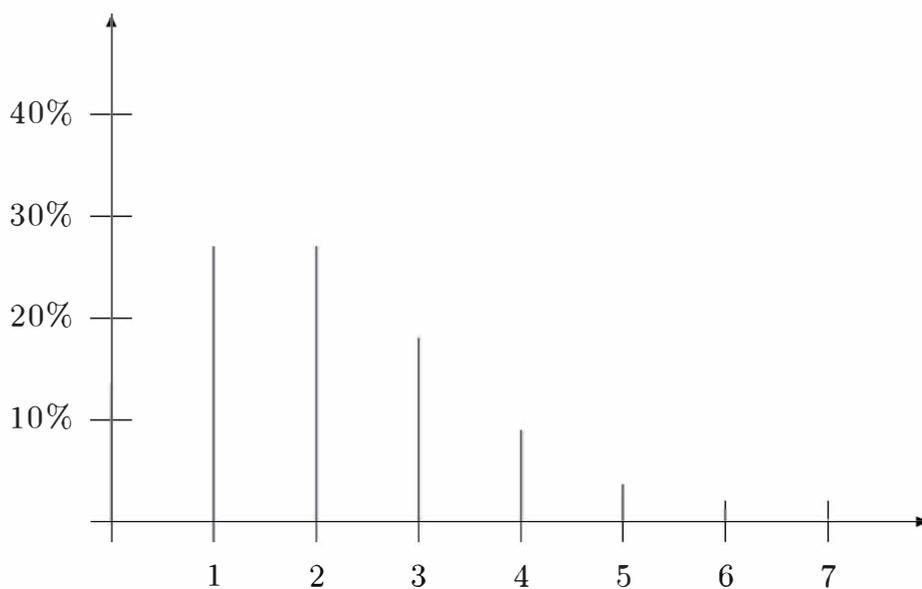
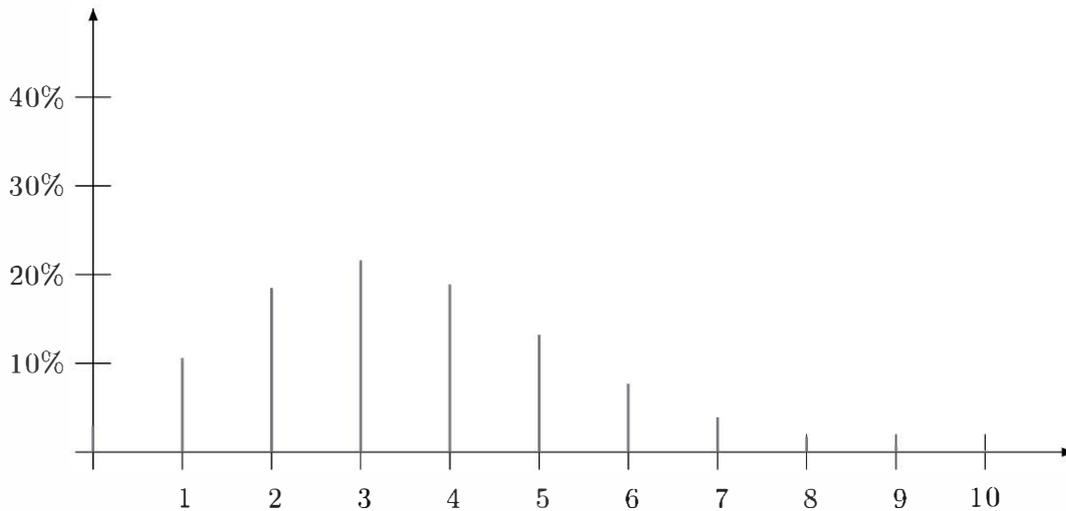


Figure 6-6 Distribution de Poisson pour $\lambda = 2$

Comparaison des observations : pour les valeurs de λ faible, la distribution est décroissante. Lorsque λ augmente, la distribution ressemble plus à une distribution normale. Cette remarque sera utile pour effectuer des approximations de lois.

Figure 6-7 Distribution de Poisson pour $\lambda = 3,5$

6.3 LOI NORMALE

a) Définition de la loi

La loi normale est une loi continue sur \mathbb{R} . Elle dépend de deux paramètres :

- m : sa moyenne (ou son espérance) ;
- σ : son écart-type.

Sa densité est :

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$$

Cette loi s'utilise dans la plupart des problèmes mettant en jeu des variables continues. On peut approximer de nombreuses lois discrètes ou continues par la loi normale.

b) Représentation graphique

La courbe représentative de la densité d'une loi normale est dite « en cloche », centrée sur la moyenne et son étalement dépend de l'écart-type.

La figure suivante est la courbe représentative de la loi normale pour laquelle $m = 0$ et $\sigma = 1$.

Remarque : cette loi est la loi normale centrée ($m = 0$) réduite ($\sigma = 1$).

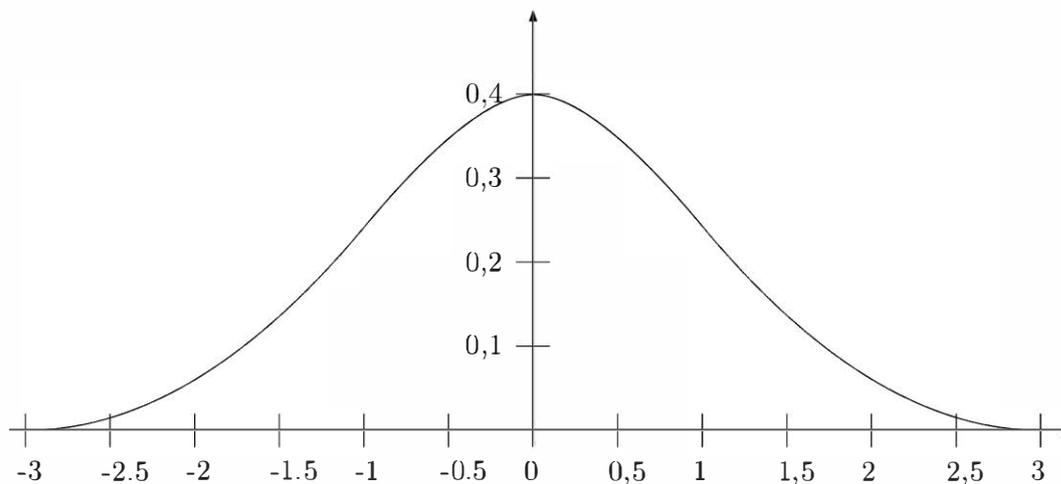


Figure 6-8 Densité de la loi normale centrée réduite

Influence de m

Pour observer l'influence de la moyenne, on fixe $\sigma = 1$ puis on choisit $m_1 < m_2$

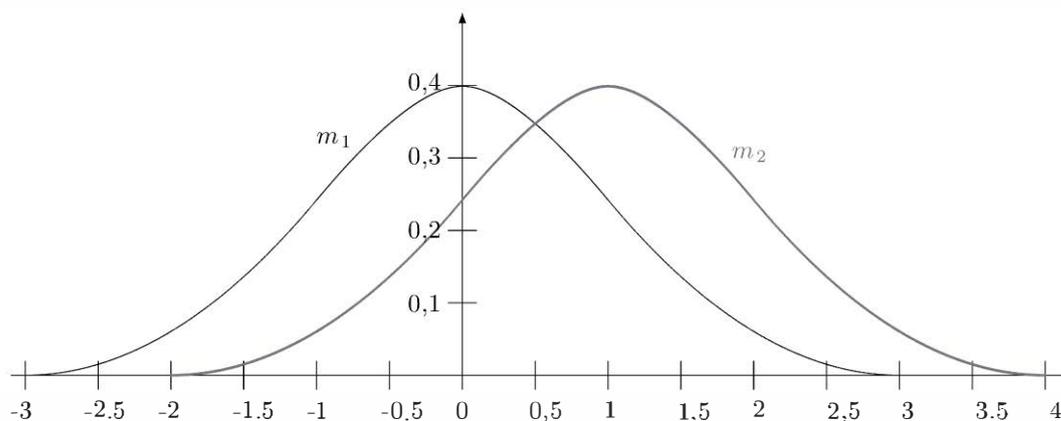


Figure 6-9 Influence de la moyenne

L'allure générale de la courbe est inchangée lorsque la moyenne varie, on constate une translation de la gauche vers la droite de la courbe lorsque la moyenne augmente (ici $m_2 > m_1$).

Cette observation se comprend car l'étalement de la courbe dépend uniquement de l'écart-type. Une variation de la moyenne ne fera varier que l'axe de symétrie de la courbe.

Influence de σ

Pour observer l'influence de l'écart-type, on fixe $m = 0$ et on choisit $\sigma_1 < \sigma_2$

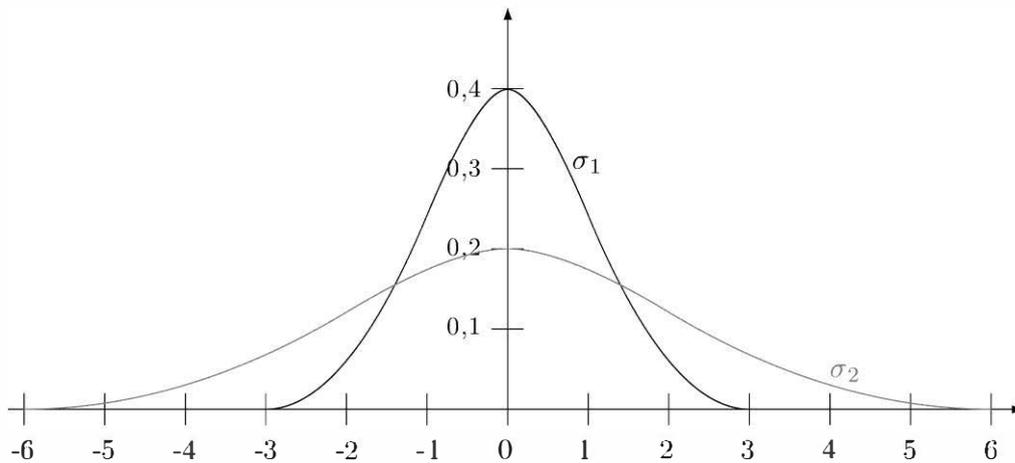


Figure 6-10 Influence de l'écart-type

L'axe de symétrie de la courbe est inchangé car la moyenne est inchangée. En revanche, l'augmentation de l'écart-type a pour effet d'écraser la courbe en augmentant ainsi son étalement.

c) Comment calculer des probabilités avec la loi normale ?

La loi normale étant une loi continue, on a $P(X = u) = 0$ quelle que soit la valeur de u . Ainsi, pour la loi normale, on a :

$$P(X \leq u) = P(X < u) + P(X = u) = P(X < u)$$



Comment calculer $P(X < u)$?

Théoriquement la fonction de répartition de la loi normale est le calcul qui permet de calculer $P(X < u)$ par la formule : $P(X < u) = \int_{-\infty}^u \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$

Le calcul de cette intégrale est trop long pour être une méthode efficace d'utilisation de la loi normale.

Pour éviter les calculs, on dispose de la table de la loi normale centrée ($m = 0$) réduite ($\sigma = 1$) : voir *Tableau 6-4*, page suivante.

Tableau 6-4 Table de la loi normale centrée réduite

u	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997

Comment utiliser la table de la loi normale centrée réduite ?

Cette table indique les valeurs de $P(X < u)$ pour u compris entre 0 et 3,9. La première colonne indique le premier chiffre après la virgule de u et la première ligne indique le second chiffre après la virgule.

- Si $0 \leq u < 4$

Par exemple pour le calcul de $P(X < 1,64)$, on cherche dans la première colonne 1,6 puis dans la première ligne 0,04. À l'intersection de cette ligne et de cette colonne, on trouve :

$$P(X < 1,64) = 0,94950$$

Pour trouver $P(X > 1,64)$, on cherche la probabilité complémentaire de $P(X < 1,64)$:

$$P(X > 1,64) = 1 - 0,94950 = 0,0505$$

- Si $-4 < u < 0$

Par exemple, pour le calcul de $P(X < -1,64)$, on se rappelle que la densité de la loi normale centrée réduite est symétrique par rapport à l'axe des ordonnées, ainsi on a la relation : $P(X < -u) = P(X > u)$

On a donc $P(X < -1,64) = P(X > 1,64) = 0,0505$

- Si $u \geq 4$

On a $P(X < u) = 1$ avec une précision de 4 chiffres après la virgule.

- Si $u \leq -4$

On a $P(X < u) = 0$ avec une précision de 4 chiffres après la virgule.

Comment utiliser cette table pour une loi normale quelconque ?

L'intérêt de cette table ne se limite pas à la loi normale centrée réduite.

La propriété suivante est utile :

Si X suit une loi normale de moyenne m et d'écart-type σ alors

$U = \frac{X - m}{\sigma}$ suit une loi normale centrée réduite.

Exemple. Si X suit une loi normale de moyenne 5 et d'écart-type 4. Calculons $P(X < 6)$.

Ce calcul est impossible directement, on va donc « centrer réduire » la variable X .

$$\text{Ainsi } P(X < 6) = P\left(\frac{X - 5}{4} < \frac{6 - 5}{4}\right) = P(U < 0,25) = 0,59871$$

Le problème inverse peut se poser ; on dispose de $P(X < u)$ et on cherche la valeur de u associée sachant que X suit une loi normale centrée réduite.

Pour cela on utilise la table de la loi normale inverse. (cf. page suivante)

Cette table se lit de deux façons différentes selon que la probabilité $P(X < u)$ soit inférieure ou supérieure à 0,5.

- Si $P(X < u) > 0,5$

La dernière colonne se lit du bas vers le haut pour donner les deux premiers chiffres de la probabilité. La dernière ligne se lit de droite à gauche et indique le troisième chiffre de la probabilité. Par exemple, si on cherche u tel que $P(X < u) = 0,76$, on cherche 0,76 dans la dernière colonne et 0,00 dans la dernière ligne. À l'intersection de la ligne et de la colonne, on trouve $u = 0,7192$

- Si $P(X < u) < 0,5$

La première colonne se lit du haut vers le bas pour donner les deux premiers chiffres de la probabilité. La première ligne se lit de gauche à droite et indique le troisième chiffre de la probabilité. Comme la courbe de la densité de la loi normale centrée réduite est symétrique par rapport à l'axe des ordonnées, la valeur de u trouvée de le tableau sera à compter négativement. Par exemple, si on cherche u tel que $P(X < u) = 0,216$, on cherche 0,21 dans la première colonne et 0,006 dans la première ligne. À l'intersection de la ligne et de la colonne, on trouve $u = -0,7858$

Remarque : pour une loi normale quelconque, la méthode de « centrer réduire » la variable s'applique dans ce cas aussi.

6.4 APPROXIMATIONS DE LOIS

La loi binomiale et la loi de Poisson ne sont pas toujours simples d'utilisation. Par exemple, le calcul des combinaisons dans la loi binomiale est très long. Considérons également une loi de Poisson de paramètre 60, si on cherche à calculer $P(X < 58)$, on va devoir calculer $P(X = 0) + P(X = 1) + P(X = 2) + \dots + P(X = 57)$, ce qui représente 58 calculs en tout. Calculer $P(X < 58)$ est bien plus simple par lecture de la table de la loi normale centrée réduite. Il est donc légitime de chercher à faire des approximations de lois. Il reste à voir quelles approximations sont envisageables et sous quelles conditions.

Tableau 6-5 Table de la loi normale inverse

P	0,000	0,001	0,002	0,003	0,004	0,005	0,006	0,007	0,008	0,009	0,010	
0,00	∞	3,0902	2,8782	2,7478	2,6521	2,5758	2,5121	2,4573	2,4089	2,3656	2,3263	0,99
0,01	2,3263	2,2904	2,2571	2,2262	2,1973	2,1701	2,1444	2,1201	2,0969	2,0749	2,0537	0,98
0,02	2,0537	2,0335	2,0141	1,9954	1,9774	1,9600	1,9421	1,9268	1,9110	1,8957	1,8808	0,97
0,03	1,8808	1,8663	1,8522	1,8384	1,8250	1,8119	1,7991	1,7866	1,7744	1,7624	1,7507	0,96
0,04	1,7507	1,7392	1,7279	1,7169	1,7060	1,6954	1,6849	1,6747	1,6646	1,6546	1,6449	0,95
0,05	1,6449	1,6352	1,6258	1,6164	1,6072	1,5982	1,5893	1,5805	1,5718	1,5632	1,5548	0,94
0,06	1,5548	1,5464	1,5382	1,5301	1,5220	1,5141	1,5063	1,4985	1,4909	1,4833	1,4758	0,93
0,07	1,4758	1,4684	1,4611	1,4538	1,4466	1,4395	1,4325	1,4255	1,4187	1,4118	1,4051	0,92
0,08	1,4051	1,3984	1,3917	1,3852	1,3787	1,3722	1,3658	1,3595	1,3532	1,3469	1,3408	0,91
0,09	1,3408	1,3346	1,3285	1,3225	1,3165	1,3106	1,3047	1,2988	1,1930	1,2873	1,2816	0,90
0,10	1,2816	1,2759	1,2702	1,2646	1,2591	1,2536	1,2481	1,2426	1,2372	1,2319	1,2265	0,89
0,11	1,2265	1,2212	1,2160	1,2107	1,2055	1,2004	1,1952	1,1901	1,1850	1,1800	1,1750	0,88
0,12	1,1750	1,1700	1,1650	1,1601	1,1552	1,1503	1,1455	1,1407	1,1359	1,1311	1,1264	0,87
0,13	1,1264	1,1217	1,1170	1,1123	1,1077	1,1031	1,0985	1,0939	1,0893	1,0848	1,0803	0,86
0,14	1,0803	1,0758	1,0714	1,0669	1,0625	1,0581	1,0537	1,0494	1,0450	1,0407	1,0364	0,85
0,15	1,0364	1,0322	1,0279	1,0237	1,0194	1,0152	1,0110	1,0069	1,0027	0,9986	0,9945	0,84
0,16	0,9945	0,9904	0,9863	0,9822	0,9782	0,9741	0,9701	0,9661	0,9621	0,9581	0,9542	0,83
0,17	0,9542	0,9502	0,9463	0,9424	0,9385	0,9346	0,9307	0,9269	0,9230	0,9192	0,9154	0,82
0,18	0,9154	0,9116	0,9078	0,9040	0,9002	0,8965	0,8927	0,8890	0,8853	0,8816	0,8779	0,81
0,19	0,8779	0,8742	0,8705	0,8669	0,8633	0,8596	0,8560	0,8524	0,8488	0,8452	0,8416	0,80
0,20	0,8416	0,8381	0,8345	0,8310	0,8274	0,8239	0,8204	0,8169	0,8134	0,8099	0,8064	0,79
0,21	0,8064	0,8030	0,7995	0,7961	0,7926	0,7892	0,7858	0,7824	0,7790	0,7756	0,7722	0,78
0,22	0,7722	0,7688	0,7655	0,7621	0,7588	0,7554	0,7521	0,7488	0,7454	0,7421	0,7388	0,77
0,23	0,7388	0,7356	0,7323	0,7290	0,7257	0,7225	0,7192	0,7160	0,7128	0,7095	0,7063	0,76
0,24	0,7063	0,7031	0,6999	0,6967	0,6935	0,6903	0,6871	0,6840	0,6808	0,6776	0,6745	0,75
0,25	0,6745	0,6713	0,6682	0,6651	0,6620	0,6588	0,6557	0,6526	0,6495	0,6464	0,6433	0,74
0,26	0,6433	0,6403	0,6372	0,6341	0,6311	0,6280	0,6250	0,6219	0,6189	0,6158	0,6128	0,73
0,27	0,6128	0,6098	0,6068	0,6038	0,6008	0,5978	0,5948	0,5918	0,5888	0,5858	0,5828	0,72
0,28	0,5828	0,5799	0,5769	0,5740	0,5710	0,5681	0,5651	0,5622	0,5592	0,5563	0,5534	0,71
0,29	0,5534	0,5505	0,5476	0,5446	0,5417	0,5388	0,5359	0,5330	0,5302	0,5273	0,5244	0,70
0,30	0,5244	0,5215	0,5187	0,5158	0,5129	0,5101	0,5072	0,5044	0,5015	0,4987	0,4959	0,69
0,31	0,4959	0,4930	0,4902	0,4874	0,4845	0,4817	0,4789	0,4761	0,4733	0,4705	0,4677	0,68
0,32	0,4677	0,4649	0,4621	0,4593	0,4565	0,4538	0,4510	0,4482	0,4454	0,4427	0,4399	0,67
0,33	0,4399	0,4372	0,4344	0,4316	0,4289	0,4261	0,4234	0,4207	0,4179	0,4152	0,4125	0,66
0,34	0,4125	0,4097	0,4070	0,4043	0,4016	0,3989	0,3961	0,3934	0,3907	0,3880	0,3853	0,65
0,35	0,3853	0,3826	0,3799	0,3772	0,3745	0,3719	0,3692	0,3665	0,3638	0,3611	0,3585	0,64
0,36	0,3585	0,3558	0,3531	0,3505	0,3478	0,3451	0,3425	0,3398	0,3372	0,3345	0,3319	0,63
0,37	0,3319	0,3292	0,3266	0,3239	0,3213	0,3186	0,3160	0,3134	0,3107	0,3081	0,3055	0,62
0,38	0,3055	0,3029	0,3002	0,2976	0,2950	0,2924	0,2898	0,2871	0,2845	0,2819	0,2793	0,61
0,39	0,2793	0,2767	0,2741	0,2715	0,2689	0,2663	0,2637	0,2611	0,2585	0,2559	0,2533	0,60
0,40	0,2533	0,2508	0,2482	0,2456	0,2430	0,2404	0,2378	0,2353	0,2327	0,2301	0,2275	0,59
0,41	0,2275	0,2250	0,2224	0,2198	0,2173	0,2147	0,2121	0,2096	0,2070	0,2045	0,2019	0,58
0,42	0,2019	0,1993	0,1968	0,1942	0,1917	0,1891	0,1866	0,1840	0,1815	0,1789	0,1764	0,57
0,43	0,1764	0,1738	0,1713	0,1687	0,1662	0,1637	0,1611	0,1586	0,1560	0,1535	0,1510	0,56
0,44	0,1510	0,1484	0,1459	0,1434	0,1408	0,1383	0,1358	0,1332	0,1307	0,1282	0,1257	0,55
0,45	0,1257	0,1231	0,1206	0,1181	0,1156	0,1130	0,1105	0,1080	0,1055	0,1030	0,1004	0,54
0,46	0,1004	0,0979	0,0954	0,0929	0,0904	0,0878	0,0853	0,0828	0,0803	0,0778	0,0753	0,53
0,47	0,0753	0,0728	0,0702	0,0677	0,0652	0,0627	0,0602	0,0577	0,0552	0,0527	0,0502	0,52
0,48	0,0502	0,0476	0,0451	0,0426	0,0401	0,0376	0,0351	0,0326	0,0301	0,0276	0,0251	0,51
0,49	0,0251	0,0226	0,0201	0,0175	0,0150	0,0125	0,0100	0,0075	0,0050	0,0025	0,0000	0,50
	0,010	0,009	0,008	0,007	0,006	0,005	0,004	0,003	0,002	0,001	0,000	P

2 Probabilités

a) Approximation d'une loi binomiale par une loi de Poisson

La loi binomiale dépend de deux paramètres n et p , alors que la loi de Poisson ne dépend que du paramètre λ . Pour qu'une loi binomiale soit au plus proche d'une loi de Poisson, on doit au moins souhaiter que ces deux lois aient la même espérance. L'espérance de la loi binomiale étant np et celle de la loi de Poisson étant λ , il faut $\lambda = np$. Cette condition nécessaire n'est pas suffisante pour réaliser une telle approximation, **théoriquement** l'approximation est parfaite lorsque :

$$\begin{cases} n & \longrightarrow & \infty \\ p & \longrightarrow & 0 \\ np & = & \text{constante} \end{cases}$$

En **pratique**, la condition :

$$\begin{cases} n & > & 30 \\ np & < & 5 \end{cases}$$

ou :

$$\begin{cases} n & > & 50 \\ p & < & 0,1 \end{cases}$$

est suffisante pour envisager l'approximation.

Remarque : la condition p très petit ($p < 0,1$) peut être en réalité remplacée par p très grand ($p > 0,9$), car si la réussite est de probabilité p dans la loi binomiale, l'échec est de probabilité $1 - p$. Par conséquent, p grand implique $1 - p$ petit et il est possible de faire une approximation sur la loi binomiale de paramètre $1 - p$ qui renseignera ensuite sur la loi binomiale de paramètre p .

Exemple. On considère une loi binomiale de paramètres $n = 35$ et $p = 0,1$. On est dans les conditions d'approximation de cette loi par une loi de Poisson de paramètre $\lambda = 0,1 \times 35 = 3,5$. Le *Tableau 6-6* page suivante donne une comparaison numérique de $P(X = k)$ pour la loi binomiale et pour la loi de Poisson.

Tableau 6-6 Approximation binomiale-Poisson

k	0	1	2	3	4	5	6	7	8	9
Binomiale (%)	2,50	9,73	18,39	22,47	19,98	13,76	7,65	3,52	1,37	0,45
Poisson (%)	3,02	10,56	18,50	21,58	18,88	13,22	7,71	3,86	1,69	0,66
Écart (%)	0,52	0,83	0,11	0,89	1,1	0,54	0,06	0,34	0,32	0,21

Les valeurs suivantes de k ne sont pas données car $P(X = k)$ est très proche de 0 dans les deux lois.

On constate que l'écart entre les deux lois ne dépasse pas 1,1 %. À 1 % près, on peut supposer que les deux lois sont équivalentes.

Remarque : dans l'approximation d'une loi binomiale par une loi de Poisson, l'ensemble des possibles se dilate de l'ensemble $\{0, 1, 2, \dots, n\}$ à l'ensemble \mathbb{N} . Il y a un passage d'un ensemble fini à un ensemble infini. Dans la loi de Poisson approximée, on doit donc avoir $P(X = k) \approx 0$ si $k > n$ pour être en cohérence avec la loi binomiale.

b) Approximation d'une loi de Poisson par une loi normale

Une loi de Poisson de paramètre λ peut être approximée par une loi normale de moyenne λ et d'écart-type $\sqrt{\lambda}$ si λ est grand. Dans la **pratique**, la condition $\lambda > 5$ est suffisante.

Comment passer dans le calcul d'une loi discrète à une loi continue ?

Le problème se pose car $P(0,3 < X < 0,7)$ est nulle pour une loi de Poisson et est non nulle pour une loi normale qui prend ses valeurs dans l'ensemble des réels. De plus, pour une loi normale $P(X = k) = 0$ quelque soit la valeur de k , alors que pour la loi de Poisson $P(X = k) \neq 0$ si k est un entier naturel.

Règles de calcul

Notons B la variable de Poisson à approximer par la variable normale N et notons f la densité de N .

- $P(B < k) \approx P(N < k)$ en calculant $P(N < k)$ à l'aide de la table de la loi normale.
- $P(B = k) \approx f(k) \approx P(k - 0,5 < N < k + 0,5)$, ainsi la densité de la loi normale permet d'approximer la loi de Poisson plus vite que l'utilisation de la table de la loi normale.

Exemple. On considère une loi de Poisson de paramètre $\lambda = 9$, on est dans les conditions d'approximation de cette loi par une loi normale de moyenne 9 et d'écart-type $\sqrt{9} = 3$. Le tableau suivant donne une comparaison numérique de $P(X = k)$ pour la loi de Poisson, pour la densité $f(k)$ de la loi normale et pour $P(k - 0,5 < X < k + 0,5)$ dans la loi normale. Les résultats dans le tableau sont en pourcentage.

Tableau 6-7 Approximation Poisson-Normale

k	0	1	2	3	4	5
Poisson	0,01	0,11	0,5	1,50	3,37	6,07
$f(k)$	0,15	0,38	0,87	1,80	3,32	5,47
$P([k - \frac{1}{2}; k + \frac{1}{2}])$	0,15	0,39	0,89	1,82	3,34	5,49

k	6	7	8	9	10	11
Poisson	9,11	11,71	13,17	13,17	11,86	9,70
$f(k)$	8,07	10,65	12,58	13,30	12,58	10,65
$P([k - \frac{1}{2}; k + \frac{1}{2}])$	8,07	10,62	12,53	13,24	12,53	10,62

On rappelle les formules dans cet exemple : $P(X = k) = e^{-9} \times \frac{9^k}{k!}$

pour la loi de Poisson et $f(x) = \frac{1}{\sqrt{2\pi}3} e^{-0,5(\frac{x-9}{3})^2}$ pour la densité de

la loi normale.

Les résultats du tableau indiquent qu'il est quasi équivalent de calculer $f(k)$

et $P\left(\left[k - \frac{1}{2}; k + \frac{1}{2}\right]\right)$, par conséquent on se contentera d'un seul

des deux calculs dans les exercices. D'autre part, l'écart entre les valeurs prises par la loi de Poisson et celles de la loi normale excède rarement 1 %. Cette observation confirme le bien-fondé de cette approximation.

c) Approximation d'une loi binomiale par une loi normale

Une loi binomiale de paramètres n et p peut être approximée par une loi normale de moyenne np et d'écart-type $\sqrt{np(1-p)}$.

Plusieurs règles existent pour justifier l'approximation par une loi normale.

Une des règles les plus simples est $np \pm 3\sqrt{np(1-p)} \in [0; n]$

Exemple. Considérons une loi binomiale de paramètres $n = 10$ et $p = 0,5$. On a $np + 3\sqrt{np(1-p)} = 9,74 < 10$

et $np - 3\sqrt{np(1-p)} = 0,26 > 0$

Ainsi l'approximation par une loi normale de moyenne 5 et d'écart-type 1,58 semble possible. Vérifions le numériquement :

Tableau 6-8 Approximation Binomiale-Normale

k	0	1	2	3	4	5	6	7	8	9	10
Binomiale	0,10	0,98	4,40	11,72	20,51	24,62	20,51	11,72	4,40	0,98	0,10
$f(k)$	0,17	1,02	4,16	11,33	20,67	25,25	20,67	11,33	4,16	1,02	0,17

On constate des résultats dont l'écart ne dépasse pas 1%. L'approximation est donc correcte.

Remarque : comme dans la partie précédente, il est équivalent de calculer $P(k - 0,5 < X < k + 0,5)$ et $f(k)$ pour la loi normale.



POINTS CLEFS

► Lois fondamentales

Lois	Ensemble des possibles	Paramètres	Définition
Loi binomiale	$\{0, 1, 2, \dots, n\}$	n et p	$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$
Loi de Poisson	\mathbb{N}	λ	$P(X = k) = e^{-\lambda} \times \frac{\lambda^k}{k!}$
Loi normale	\mathbb{R}	m et σ	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$

► Espérance et variance

Lois	Espérance	Variance	Écart-type
Loi binomiale	np	$np(1-p)$	$\sqrt{np(1-p)}$
Loi de Poisson	λ	λ	$\sqrt{\lambda}$
Loi normale	m	σ^2	σ

➤ **Mode de calcul des probabilités**

- Pour la loi binomiale et la loi de Poisson, la formule $P(X = k)$ permet de réaliser l'ensemble des calculs.
- Pour la loi normale, la table de la loi normale centrée réduite permet de calculer simplement les probabilités du type $P(X < u)$. Pour cela, il faut centrer réduire la variable X par la formule : $\frac{X-m}{\sigma} = U$
- La table inverse sert à retrouver u lorsque $P(X < u)$ est connu.

➤ **Approximations de lois**

Loi à approximer	Approximation	Conditions
Binomiale (n, p)	Poisson(np)	$\{n > 30 \text{ et } np < 5\}$ ou $\{n > 50 \text{ et } p < 0,1\}$
Binomiale (n, p)	Normale $(np, \sqrt{np(1-p)})$	$np \pm 3\sqrt{np(1-p)} \in [0; n]$
Poisson (λ)	Normale $(\lambda, \sqrt{\lambda})$	$\lambda > 5$

EXERCICES

6.1 Questionnaire

Un questionnaire comporte dix questions. Pour chacune, il y a quatre réponses possibles. Un individu répond au hasard à toutes ces questions.

- a) Quelle est la probabilité qu'il réponde juste à toutes les questions ?
- b) Quelle est la probabilité qu'il réponde juste à 5 questions ?
- c) Quelle est la probabilité qu'il réponde juste à au moins 8 questions ?
- d) En moyenne, à combien de questions va-t-il répondre juste ?

6.2 Probabilité de panne

Une entreprise dispose d'une machine dont la probabilité de tomber en panne en une journée est de $1/500$.

- a) Quelle est la probabilité que cette machine tombe en panne au moins une fois en 20 jours ?
- b) Quelle est la probabilité que cette machine tombe strictement plus d'une fois en panne sur cette période ?

6.3 Ouverture d'une nouvelle caisse

Un magasin possède 4 caisses. Le nombre de clients qui attendent à chacune de ces caisses suit une loi de Poisson de paramètre 3,2. Un nombre de clients supérieur ou égal à 4 par caisse risque de ne pas plaire à la clientèle. Il faudrait alors ouvrir une nouvelle caisse.

Quelle est la probabilité d'ouvrir une nouvelle caisse ?

6.4 Ordinateurs

On réalise une enquête sur 200 personnes qui indiquent le nombre d'ordinateurs chez eux :

Nombre d'ordinateurs	0	1	2	3	4	5	6 et plus
Nombre de personnes	60	73	43	17	5	2	0

- Déterminer le nombre moyen d'ordinateurs par personne.
- Déterminer la variance du nombre d'ordinateurs par personne.
- Quelle loi permettrait de modéliser cette statistique ?
- Vérifier par un calcul d'écart votre proposition.

6.5 Utilisation des tables

Soit X une variable qui suit une loi normale centrée réduite.

- Calculer $P(X < 0)$, $P(X < 1,34)$, $P(1 < X < 2)$, $P(X > 3,45)$ et $P(X < -2,67)$.
- Trouver u tel que $P(X < u) = 0,75$, $P(X < u) = 0,23$ et $P(-u < X < u) = 0,1$.

6.6 Résultats à un examen

La moyenne des notes à un examen suit une loi normale de moyenne 9,5 et d'écart-type 4,5. Les étudiants ayant une moyenne inférieure à 7 ne sont pas admissibles. Les étudiants ayant une moyenne entre 7 et 9 doivent repasser une des épreuves pour être admissible. Les étudiants ayant entre 9 et 13 sont admissibles et doivent passer un oral. Les étudiants ayant plus de 13 sont directement admis.

Sur un ensemble de 400 000 étudiants qui se présentent à cet examen, déterminer une estimation du nombre d'étudiants de chaque catégorie.

6.7 Bénéfices

Le seuil de rentabilité est atteint lorsque le bénéfice est positif. Une entreprise fabrique un produit unique. Le coût variable unitaire est éva-

lué à 35 euros. Le prix de vente unitaire est 58 euros. Les charges de structure de la période sont estimées à 50 000 euros. Les quantités vendues suivent une loi normale de moyenne 4 000 unités et oscillent entre 3 500 et 4 500 unités dans 80 % des cas.

- a) Calculer l'écart-type des quantités vendues.
- b) Calculer la probabilité d'atteindre le seuil de rentabilité.
- c) Calculer la probabilité d'atteindre un bénéfice supérieur à 30 000 euros.

6.8 Pièces cylindriques

Une entreprise produit des pièces cylindriques dont la caractéristique essentielle est le diamètre. La mesure de ce diamètre est reconnue comme étant une variable normale. Sur un ensemble de 100 000 pièces qui viennent d'être produites, on en dénombre 21 300 dont le diamètre est inférieur à 1,21 mm et 5 600 dont le diamètre est supérieur à 1,32 mm.

- a) Déterminer la moyenne et l'écart-type du diamètre des pièces produites.
- b) Déterminer la médiane et les quartiles du diamètre des pièces.

6.9 Salaires

Le salaire mensuel d'un commercial suit une loi normale de moyenne 3 000 euros et d'écart-type 400 euros.

- a) Déterminer la probabilité que le salaire trimestriel d'un commercial dépasse 10 000 euros.
- b) On s'intéresse à l'écart mensuel de salaire entre deux commerciaux. Quelle est la probabilité que celui-ci dépasse 500 euros ?

6.10 Voyage en avion

Un avion décolle à 5h25 du matin. On note que 10 % des passagers arrivent avant l'heure d'enregistrement des bagages et 5 % d'entre eux arrivent après la fermeture de l'enregistrement et ne peuvent pas embarquer. L'heure d'ouverture de l'enregistrement est 3h50 et l'heure de fermeture de l'enregistrement est 5h.

- a) Déterminer les paramètres de la loi normale qui représente l'heure d'arrivée d'un passager.
- b) Combien la compagnie doit-elle enregistrer de réservations pour s'assurer qu'un avion de 200 places soit complet ?
- c) Que se passe-t-il si l'avion a 10 minutes de retard et que la fermeture de l'enregistrement prend aussi 10 minutes de retard ?

6.11 loi de Poisson

Le nombre de personnes à une caisse suit une loi de Poisson de moyenne 12.

Quelle est la probabilité de trouver plus de 20 personnes à cette caisse ?

6.12 Production d'une usine

Une usine a produit des pièces dont 2 % ont un défaut. On teste un ensemble de 10 000 pièces.

- a) Que dire de l'utilisation de la loi Binomiale pour cette étude ?
 b) À l'aide d'une approximation de loi, déterminer la probabilité de trouver plus de 2,2 % de pièces défectueuses dans l'ensemble des 10 000 pièces.

SOLUTIONS**Exercice 6.1**

Le nombre de réponses justes au questionnaire suit une loi binomiale de paramètres $n = 10$, car il y a dix questions, et $p = 0,25$, car pour chaque question il y a une chance sur quatre de répondre juste. Notons X le nombre de réponses justes.

a) $P(X = 10) = 0,25^{10} = 9,53 \times 10^{-7}$. La probabilité de répondre juste à toutes les questions est donc négligeable.

$$\begin{aligned} \text{b) } P(X = 5) &= \frac{10!}{5! \times (10 - 5)!} 0,25^5 0,75^5 \\ &= 252 \times 0,000977 \times 0,237305 = 5,84 \% \end{aligned}$$

$$\begin{aligned} \text{c) } P(X \geq 8) &= P(X = 10) + P(X = 9) + P(X = 8) \\ &= 9,53 \times 10^{-7} + 9 \times 0,25^9 \times 0,75 \\ &\quad + \frac{10!}{8! \times 2!} 0,25^8 \times 0,75^2 = 0,0413 \% \end{aligned}$$

d) Il s'agit de l'espérance de la loi binomiale :

$$10 \times 0,25 = 2,5 \text{ réponses justes en moyenne.}$$

Exercice 6.2

a) Le nombre de pannes en 20 jours, X , suit une loi binomiale de paramètres $n = 20$ et $p = 1/500$. On cherche $P(X \geq 1)$. Le calcul direct est trop long, on utilise donc :

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \left(\frac{499}{500}\right)^{20} = 3,925 \%$$

$$\begin{aligned} \text{b) De même } P(X > 1) &= 1 - P(X = 0) - P(X = 1) \\ &= 0,03925 - 20 \times \frac{1}{500} \times \left(\frac{499}{500}\right)^{19} = 0,074 \% \end{aligned}$$

Exercice 6.3

Déterminons la probabilité d'avoir un nombre de clients supérieur ou égal à 4 à une caisse donnée. Notons X le nombre de clients à une caisse. X suit une loi de Poisson de paramètre 3,2.

On a :

$$\begin{aligned} P(X \geq 4) &= 1 - P(X = 0) - P(X = 1) - P(X = 2) - P(X = 3) \\ &= 1 - e^{-3,2} \times \frac{3,2^0}{0!} - e^{-3,2} \times \frac{3,2^1}{1!} - e^{-3,2} \times \frac{3,2^2}{2!} - e^{-3,2} \times \frac{3,2^3}{3!} \\ &= 39,75 \% \end{aligned}$$

Il y a 4 caisses dans ce magasin, on décide d'ouvrir une caisse supplémentaire seulement si les quatre caisses ont plus de 4 clients. Notons Y , le nombre de caisses ayant plus de quatre clients, Y suit une loi binomiale de paramètres $n = 4$ car il y a quatre caisses et $p = 39,75\%$. $P(Y = 4) = 0,3975^4 = 2,5\%$. Ainsi il y a 2,5 % de chances d'ouvrir une caisse supplémentaire.

Exercice 6.4

a) Par la formule de la moyenne pondérée en statistiques, on trouve un nombre moyen d'ordinateurs égal à

$$\frac{0 \times 60 + 1 \times 73 + 2 \times 43 + 3 \times 17 + 4 \times 5 + 5 \times 2}{200} = 1,2$$

b) La variance de cette série est

$$\begin{aligned} &\frac{0^2 \times 60 + 1^2 \times 73 + 2^2 \times 43 + 3^2 \times 17 + 4^2 \times 5 + 5^2 \times 2}{200} - 1,2^2 \\ &= 1,2 \end{aligned}$$

c) On constate que la variance est égale à la moyenne de cette série, le nombre d'ordinateurs par personnes pourrait être modélisé par une loi de Poisson de moyenne 1,2.

d) Le tableau suivant présente les valeurs de la loi de Poisson, les proportions de l'échantillon et le calcul de l'écart entre les deux.

On constate des écarts très faibles entre le modèle et la statistique. Il est donc pertinent de modéliser cette observation par une loi de Poisson.

Nombre d'ordinateurs	0	1	2	3	4	5	6	7
Loi de Poisson	30,12 %	36,14 %	21,69 %	8,67 %	2,60 %	0,63 %	0,125 %	0,02 %
Proportions	30 %	36,5 %	21,5 %	8,5 %	2,5 %	1 %	0 %	0 %
Écart	0,12 %	0,36 %	0,19 %	0,17 %	0,1 %	0,37 %	0,125 %	0,02 %

Exercice 6.5

a)

- $P(X < 0) = 50\%$
- $P(X < 1,34) = 90,988\%$
- $P(1 < X < 2) = P(X < 2) - P(X < 1) = 0,97725 - 0,84134 = 13,591\%$
- $P(X > 3,45) = 1 - P(X < 3,45) = 1 - 0,99972 = 0,028\%$
- $P(X < -2,67) = P(X > 2,67) = 1 - P(X < 2,67) = 1 - 0,99621 = 0,379\%$

b)

- $P(X < u) = 0,75 \Leftrightarrow u = 0,6745$
- $P(X < u) = 0,23 \Leftrightarrow u = -0,7388$
- $P(-u < X < u) = 0,1 \Leftrightarrow P(X < u) - P(X < -u) = 0,1$
 $\Leftrightarrow P(X < u) - P(X > u) = 0,1$
 $\Leftrightarrow P(X < u) - (1 - P(X < u)) = 0,1$
 $\Leftrightarrow 2P(X < u) - 1 = 0,1$
 $\Leftrightarrow P(X < u) = 0,55$
 $\Leftrightarrow u = 0,1257$

Exercice 6.6

À l'aide de la table de la loi normale, nous allons dans un premier temps calculer la probabilité d'appartenir à un des intervalles de notes considérés dans l'énoncé, ensuite il sera possible de faire une estimation du nombre d'étudiants concernés par chaque intervalle.

– **Proportion d'étudiants non admissibles** : il s'agit de ceux ayant une note inférieure à 7. Déterminons $P(X < 7)$ où X représente une variable normale de moyenne 9,5 et d'écart-type 4,5.

À l'aide de la table, on va « centrer réduire » la variable :

$$\begin{aligned} P(X < 7) &= P\left(\frac{X - 9,5}{4,5} < \frac{7 - 9,5}{4,5}\right) \\ &= P(U < -0,56) = P(U > 0,56) = 1 - P(U < 0,56) \\ &= 1 - 0,71226 = 0,28774 \end{aligned}$$

Ainsi on peut estimer le nombre d'étudiants concernés à :

$$400\,000 \times 0,28774 = 115\,096$$

– **Proportion d'étudiants entre 7 et 9 :**

$$\begin{aligned} P(7 < X < 9) &= P(X < 9) - P(X < 7) \\ &= P\left(U < \frac{9 - 9,5}{4,5}\right) - 0,28774 \\ &= P(U < -0,11) - 0,28774 = P(U > 0,11) - 0,28774 \\ &= 1 - P(U < 0,11) - 0,28774 \\ &= 1 - 0,5438 - 0,28774 = 0,16846 \end{aligned}$$

Le nombre d'étudiants concernés est estimé à :

$$400\,000 \times 0,16846 = 67\,384$$

– **Proportion d'étudiants entre 9 et 13 :**

$$\begin{aligned} P(9 < X < 13) &= P(X < 13) - P(X < 9) \\ &= P\left(U < \frac{13 - 9,5}{0,5}\right) - P\left(U < \frac{9 - 9,5}{0,5}\right) \\ &= P(U < 0,78) - P(U < -0,11) \\ &= 0,7823 - (1 - 0,5438) = 0,3261 \end{aligned}$$

Le nombre d'étudiants concernés est donc $400\,000 \times 0,3261 = 130\,440$

– **Étudiants à plus de 13 :** il s'agit du reste :

$$400\,000 - 115\,096 - 67\,384 - 130\,440 = 87\,080$$

Remarque : les résultats de cet exercice sont donnés à un étudiant près, il est important de noter que cette précision n'est pas indispensable dans la mesure où il ne s'agit que d'une estimation.

Exercice 6.7

a) Dans ce problème, on dispose d'une observation mais pas de l'écart-type de la loi normale. On note Q la variable normale « quantité vendue ».

On a :

$$\begin{aligned}
 P(3\,500 < Q < 4\,500) &= P(Q < 4\,500) - P(Q < 3\,500) \\
 &= P\left(\frac{Q - 4\,000}{\sigma} < \frac{4\,500 - 4\,000}{\sigma}\right) \\
 &\quad - P\left(\frac{Q - 4\,000}{\sigma} < \frac{3\,500 - 4\,000}{\sigma}\right) \\
 &= P\left(U < \frac{500}{\sigma}\right) - P\left(U < \frac{-500}{\sigma}\right) \\
 &= P\left(U < \frac{500}{\sigma}\right) - P\left(U > \frac{500}{\sigma}\right) \\
 &= P\left(U < \frac{500}{\sigma}\right) - (1 - P\left(U < \frac{500}{\sigma}\right)) \\
 &= 2 \times P\left(U < \frac{500}{\sigma}\right) - 1
 \end{aligned}$$

Or $P(3\,500 < Q < 4\,500) = 0,8$ d'où $2 \times P\left(U < \frac{500}{\sigma}\right) - 1 = 0,8$

ainsi $P\left(U < \frac{500}{\sigma}\right) = 0,9$. À l'aide de la table inverse, on trouve :

$$\frac{500}{\sigma} = 1,2816, \text{ ainsi } \sigma \approx 390.$$

b) Notons B le bénéfice, on a $B = (58 - 35)Q - 50\,000 = 23Q - 50\,000$

On cherche :

$$\begin{aligned}
 P(B > 0) &= P(23Q - 50\,000 > 0) = P(23Q > 50\,000) \\
 &= P(Q > 2\,173,91) = P\left(U > \frac{2\,173,91 - 4\,000}{390}\right) \\
 &= P(U > -4,68) = P(U < 4,68) \approx 1
 \end{aligned}$$

Ainsi il est quasiment certain d'atteindre le seuil de rentabilité.

$$\begin{aligned}
 \text{c) } P(B > 30\,000) &= P(23Q - 50\,000 > 30\,000) = P(Q > 3\,478,26) \\
 &= P\left(U > \frac{3\,478,26 - 4\,000}{390}\right) \\
 &= P(U > -1,34) = P(U < 1,34) \approx 90,988\%
 \end{aligned}$$

Exercice 6.8

a) On ne dispose dans cet exercice ni de la moyenne ni de l'écart-type de la loi normale considérée. Notons cette variable D .

On a
$$P(D < 1,21) = \frac{21\,300}{100\,000} = 0,213$$

$$\text{Ainsi } P\left(\frac{D - m}{\sigma} < \frac{1,21 - m}{\sigma}\right) = 0,213$$

$$\text{D'où } P\left(U < \frac{1,21 - m}{\sigma}\right) = 0,213$$

$$\text{D'après la table inverse, on trouve } \frac{1,21 - m}{\sigma} = -0,7961$$

$$\text{De plus, on a } P(D > 1,32) = \frac{5\,600}{100\,000} = 0,056.$$

$$\text{Ainsi } 0,056 = P\left(U > \frac{1,32 - m}{\sigma}\right) = P\left(U < -\frac{1,32 - m}{\sigma}\right)$$

$$\text{D'où } -\frac{1,32 - m}{\sigma} = -1,5893$$

$$\text{On a donc } \begin{cases} 1,21 - m &= -0,7961\sigma \\ 1,32 - m &= 1,5893\sigma \end{cases}$$

En soustrayant la ligne 1 à la ligne 2, on trouve :

$$1,32 - 1,21 = (1,5893 + 0,7961)\sigma$$

d'où $\sigma = 0,04611$ mm et $m = 1,32 - 1,5893 \times 0,04611 = 1,2467$ mm

b) La médiane vérifie $P(D < \text{Médiane}) = 0,5$; ainsi dans le cas d'une loi normale, on a une égalité entre la moyenne et la médiane. La médiane est donc 1,2467 mm. Le premier quartile Q_1 vérifie $P(D < Q_1) = 0,25$

Ainsi $P\left(U < \frac{Q_1 - 1,2467}{0,04611}\right) = 0,25$. D'après la table inverse, on

trouve $\frac{Q_1 - 1,2467}{0,04611} = -0,6745$, ainsi $Q_1 = 1,2156$ mm.

La loi normale étant symétrique par rapport à la moyenne, le troisième quartile Q_3 va se trouver à la même distance de la médiane que Q_1 mais après la médiane. Ainsi $Q_3 = 1,2467 + (1,2467 - 1,2156) = 1,2778$ mm

Exercice 6.9

a) Pour répondre à cette question, il est nécessaire de supposer que les salaires mensuels du commercial sont indépendants. Ainsi on peut s'intéresser au salaire trimestriel du commercial. Notons M_1 le salaire mensuel du mois 1, M_2 le salaire mensuel du mois 2 et M_3 le salaire mensuel du mois 3 et T le salaire trimestriel, on a $T = M_1 + M_2 + M_3$ comme il y a trois mois dans un trimestre. Ainsi T suit une loi normale de moyenne :

$$E(T) = E(M_1 + M_2 + M_3) = E(M_1) + E(M_2) + E(M_3) = 3 \times 3\,000 = 9\,000$$

et de variance :

$$V(T) = V(M_1 + M_2 + M_3) = 3 \times 400^2 = 480\,000$$

Ainsi l'écart-type de T est $\sigma = \sqrt{480\,000} = 692,82$

On cherche ensuite :

$$\begin{aligned} P(T > 10\,000) &= P\left(\frac{T - 9\,000}{692,82} > \frac{10\,000 - 9\,000}{692,82}\right) \\ &= P(U > 1,44) = 1 - P(U < 1,44) = 1 - 0,92507 = 7,49\% \end{aligned}$$

Remarque : une erreur à ne pas faire serait de supposer que $T = 3M$ ce qui imposerait un même salaire durant les 3 mois. L'énoncé indique au contraire un salaire variable.

b) Notons M_1 le salaire mensuel du premier commercial et M_2 le salaire du second commercial. L'écart entre les deux salaires est $|M_1 - M_2|$. Comme M_1 et M_2 suivent la même loi, on cherche donc :

$$\begin{aligned} P(|M_1 - M_2| > 500) &= P(M_1 - M_2 > 500) + P(M_2 - M_1 > 500) \\ &= 2 \times P(M_1 - M_2 > 500). \end{aligned}$$

Posons $X = M_1 - M_2$, on a :

$$E(X) = E(M_1 - M_2) = E(M_1) - E(M_2) = 3\,000 - 3\,000 = 0$$

$$\text{et } V(M_1 - M_2) = V(M_1) + V(M_2) = 400^2 + 400^2 = 320\,000$$

Ainsi l'écart-type de X est $\sqrt{320\,000} = 565,69$

$$\begin{aligned} \text{On a donc } P(X > 500) &= P\left(U > \frac{500 - 0}{565,69}\right) = P(U > 0,88) \\ &= 1 - P(U < 0,88) = 1 - 0,81057 = 0,18943 \end{aligned}$$

$$\text{On a donc } P(|M_1 - M_2| > 500) = 2 \times 0,18943 = 37,88\%$$

Exercice 6.10

a) Notons X l'heure d'arrivée d'un passager. Il est nécessaire d'adopter une unique unité de mesure pour X , soit la minute, soit l'heure. Choisissons la minute : ainsi $3\text{h}50 = 230$ minutes. On a :

$$P(X < 230) = 0,1 \text{ d'où } P\left(U < \frac{230 - m}{\sigma}\right) = 0,1$$

$$\text{Ainsi } \frac{230 - m}{\sigma} = -1,2816$$

De même $5\text{h} = 300$ minutes, ainsi $P(X > 300) = 0,05$

$$\text{D'où } 1 - P\left(U < \frac{300 - m}{\sigma}\right) = 0,05$$

$$\text{Soit } P\left(U < \frac{300 - m}{\sigma}\right) = 0,95$$

$$\text{Ainsi} \quad \frac{300 - m}{\sigma} = 1,6449$$

$$\text{On a donc} \quad \begin{cases} 230 - m = -1,2816\sigma \\ 300 - m = 1,6449\sigma \end{cases}$$

$$\text{Ainsi} \quad \sigma = \frac{300 - 230}{1,6449 + 1,2816} = 23,91$$

$$\text{et} \quad m = 230 + 1,2816 \times 23,91 = 260,66$$

b) On a 5 % des passagers qui arrivent en retard. Si l'avion de 200 places doit être complet, il représente 95 % des réservations. Il est donc nécessaire de faire $\frac{200}{0,95} = 210,52$, soit 211 réservations pour être assuré de remplir l'avion.

c) La question devient : quelle est la probabilité de rater l'avion ? Cette probabilité est donc $P(X > 310)$ et non plus $P(X > 300)$. On a :

$$\begin{aligned} P(X > 310) &= P\left(U > \frac{310 - 260,66}{23,91}\right) = P(U > 2,06) \\ &= 1 - P(U < 2,06) = 1 - 0,9803 = 1,97\% \end{aligned}$$

Ainsi sur 211 réservations, il va y avoir seulement 4 passagers en retard. Par conséquent 207 passagers vont pouvoir monter dans un avion de 200 places ; ce qui est impossible et va imposer à la compagnie de donner une compensation à ces 7 passagers.

Exercice 6.11

Le calcul direct par une loi de Poisson est trop long, on a ici $\lambda = 12 > 5$, on peut donc envisager une approximation de cette loi par une loi normale de moyenne 12 et d'écart-type $\sqrt{12} \approx 3,46$. On a donc :

$$\begin{aligned} P(X > 20) &= P\left(U > \frac{20 - 12}{3,46}\right) = P(U > 2,31) \\ &= 1 - P(U < 2,31) = 1 - 0,98956 = 1,044\% \end{aligned}$$

Exercice 6.12

a) La loi binomiale est parfaitement adaptée pour cette étude car chaque pièce a exactement 2 % de chance d'avoir un défaut et on dispose d'un ensemble fini de pièces. Cependant son utilisation va conduire à des calculs trop long.

b) On a $n = 10\,000$ et $p = 0,02$, vérifions les conditions d'approximation par la loi normale de moyenne $10\,000 \times 0,02 = 200$ et d'écart-type $\sqrt{10\,000 \times 0,02 \times (1 - 0,02)} = 14$.

On a $np - 3\sqrt{np(1-p)} = 200 - 3 \times 14 = 158 > 0$

et $np + 3\sqrt{np(1-p)} = 242 < 10\,000$

Ainsi les conditions sont bien vérifiées. Dans un ensemble de 10 000 pièces, 2,2 % de pièces défectueuses représentent 220 pièces. On cherche donc :

$$\begin{aligned} P(X > 500) &= P\left(U > \frac{220 - 200}{14}\right) = P(U > 1,42857) \\ &= 1 - P(U < 1,42857) = 1 - 0,92364 = 7,64\% \end{aligned}$$

CHAPITRE 7

Estimateurs et tests d'hypothèses

OBJECTIFS

- Maîtriser la notion d'échantillon
- Connaître les estimateurs de moyenne et de proportion
- Savoir associer à un risque α le coefficient t correspondant
- Donner des estimations ponctuelles de moyenne ou de proportion
- Construire des intervalles de confiance pour la moyenne comme pour la proportion
- Connaître les formules d'intervalles de confiance pour les différences de moyenne et de proportion
- Énoncer un test d'hypothèse unilatéral ou bilatéral
- Trouver une probabilité critique et une valeur critique de validité de l'hypothèse nulle
- Conclure quant à la pertinence d'une hypothèse

PLAN

- 7.1 Échantillons
- 7.2 Estimation d'une moyenne
- 7.3 Estimation de proportion
- 7.4 Différences de moyennes ou de proportions
- 7.5 Tests d'hypothèses

7.1 ÉCHANTILLONS

Les sondages et les enquêtes d'opinions servent à extrapoler des observations sur une population. Comment un sondage sur 1 000 personnes donne-t-il une idée du résultat d'une élection ? Comment les tests de plusieurs voitures peuvent-ils assurer la fiabilité de l'ensemble de la production ? En résumé, comment des données sur une petite partie de la

population peuvent-elles renseigner sur la population entière ? La question est importante car il est généralement impossible de tester l'ensemble d'une population.

L'idée est d'extraire un **échantillon** pour cerner au mieux l'ensemble de la population. On distingue deux types d'échantillons :

- **les échantillons aléatoires** pour lesquels le choix des individus est le fait du hasard ;
- **les échantillons représentatifs** pour lesquels on retient une liste de critères représentatifs de la population. Ce choix de critères peut être la classe sociale, l'âge, le sexe ou le lieu de résidence pour une population humaine. On cherche à reproduire dans l'échantillon les catégories de la population.

Aucun de ces deux modes de constitution d'un échantillon n'est idéal.

Le **hasard** n'est jamais parfait : selon le moment de constitution de l'échantillon, une catégorie risque d'être plus représentée qu'une autre. Par exemple, si l'échantillon est réalisé dans la journée par des appels chez les individus, les personnes qui ne travaillent pas risquent d'être surreprésentées par rapport à celles qui sont au travail.

La **grille d'analyse de constitution de l'échantillon représentatif** est arbitraire et n'inclut jamais l'ensemble des critères existant, elle peut ainsi induire un préjugé d'observation.

Le **taille de l'échantillon** est un critère important pour l'analyse des résultats : plus la taille de l'échantillon est importante, plus les résultats sont fiables. Pour autant, il est coûteux de réaliser des enquêtes sur des échantillons de taille trop importante, il y a donc une taille minimale d'échantillon à rechercher pour avoir suffisamment de précisions au niveau de l'analyse.

La **taille de la population de référence** importe aussi. En effet, si la taille de la population est faible, la façon de constituer l'échantillon va influencer sur les résultats. Si on réalise l'échantillon par tirages successifs avec remises, les résultats vont différer de la réalisation d'un échantillon par tirages successifs sans remise. Dans un tirage avec remises, la probabilité d'apparition d'un critère ne varie pas d'un tirage à l'autre alors que, dans un tirage sans remise, cette probabilité évolue d'un tirage au suivant.

Ces différents points sont à retenir pour la pertinence de l'interprétation des résultats. Néanmoins, pour mettre en place des critères de choix

utiles nous allons retenir plusieurs hypothèses simplificatrices dans ce chapitre :

- les échantillons sont aléatoires et de taille importante ;
- la population de référence est de taille très grande.

Ainsi, on écarte les questions du mode de constitution de l'échantillon et du choix de sa taille.

7.2 ESTIMATION D'UNE MOYENNE

a) Construction de l'estimateur de moyenne

On cherche dans cette partie à estimer la moyenne d'un paramètre dans une population, ce paramètre peut être le diamètre d'un objet dans une production ou la durée de vie moyenne d'un composant.

On suppose que chaque élément de l'échantillon de taille n est représenté par une variable X_i et on suppose de plus que chacune des variables X_i suit une même loi normale de moyenne m et d'écart-type σ . Pour avoir une idée de la moyenne de la population, il semble naturel de s'intéresser dans un premier temps à la moyenne de l'échantillon.

$$\text{On pose alors } \bar{X} = \frac{X_1 + X_2 + X_3 + \cdots + X_n}{n}$$

\bar{X} est l'**estimateur de la moyenne** dans un échantillon. \bar{X} suit une loi normale. Cherchons à en déterminer les paramètres :

$$\begin{aligned} \text{Espérance : } E(\bar{X}) &= E\left(\frac{X_1 + X_2 + X_3 + \cdots + X_n}{n}\right) \\ &= \frac{E(X_1) + E(X_2) + E(X_3) + \cdots + E(X_n)}{n} \\ &= \frac{m + m + m + \cdots + m}{n} = \frac{n \times m}{n} = m \end{aligned}$$

$$\begin{aligned} \text{Variance : } V(\bar{X}) &= V\left(\frac{X_1 + X_2 + X_3 + \cdots + X_n}{n}\right) \\ &= \frac{V(X_1) + V(X_2) + V(X_3) + \cdots + V(X_n)}{n^2} \\ &= \frac{\sigma^2 + \sigma^2 + \sigma^2 + \cdots + \sigma^2}{n^2} = \frac{n \times \sigma^2}{n^2} = \frac{\sigma^2}{n} \end{aligned}$$

$$\text{Écart-type : } \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Conclusion : \bar{X} suit une loi normale de moyenne m et d'écart-type $\frac{\sigma}{\sqrt{n}}$

Ainsi, une estimation **ponctuelle** de la moyenne est l'espérance de \bar{X} .

Remarque : l'écart-type de \bar{X} diminue quand n augmente, ce qui indique que plus la taille de l'échantillon est grande, plus la précision est grande sur l'estimation de la moyenne.

b) Intervalle de confiance

La donnée ponctuelle d'une moyenne d'échantillon n'est pas suffisante pour tirer une conclusion sur la moyenne de la population. Un intervalle de confiance est un intervalle centré autour de m dans lequel on a un pourcentage de chance (la confiance) de trouver la moyenne de la population. On souhaite que ce pourcentage soit le plus important possible.

On note α le risque et $1 - \alpha$ la confiance. On sait que $U = \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}}$ suit

une loi normale centrée réduite, on va donc dans un premier temps chercher un réel positif t tel que $P(-t < U < t) = 1 - \alpha$.

La *Figure 7-1* représente la loi normale centrée réduite avec une valeur de t qui permet d'avoir $P(-t < U < t) = 1 - \alpha$. Par symétrie de la courbe, on en déduit que le risque α est séparé en deux à gauche et à droite de l'intervalle.

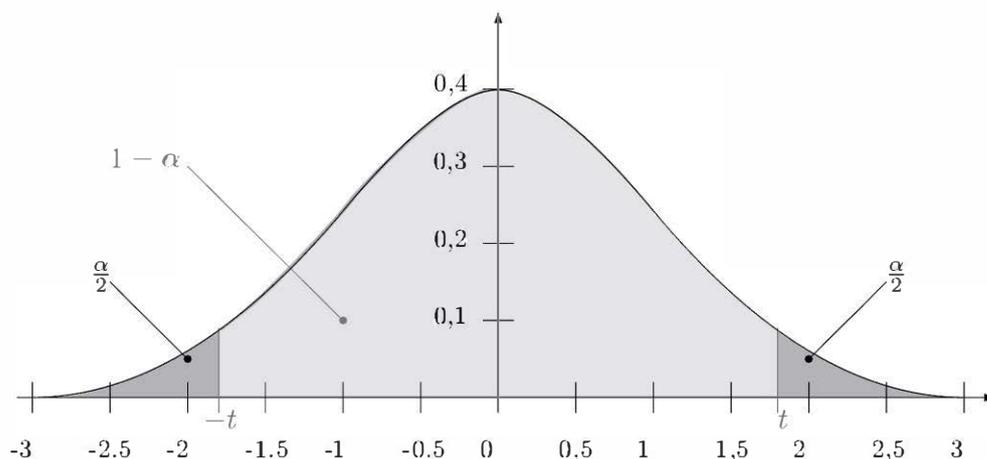


Figure 7-1 Intervalle de confiance

On a :

$$\begin{aligned} P(-t < U < t) &= P(U < t) - P(U < -t) = P(U < t) - P(U > t) \\ &= P(U < t) - (1 - P(U < t)) = 2P(U < t) - 1 \end{aligned}$$

$$\text{Ainsi } 2P(U < t) - 1 = 1 - \alpha$$

$$\text{Soit } 2P(U < t) = 2 - \alpha$$

$$\text{D'où } P(U < t) = 1 - \frac{\alpha}{2}$$

Cette dernière relation va permettre de trouver les valeurs de t associées à un risque choisi à l'aide de la table de loi normale inverse. Dans la pratique, les risques envisagés sont généralement 10 %, 5 %, 2 % ou 1 %. Le *Tableau 7-1* indique les valeurs les plus souvent rencontrées pour t :

Tableau 7-1 Valeurs de t classiques

α	$1 - \frac{\alpha}{2}$	t
10 %	0,95	1,6449
5 %	0,975	1,96
2 %	0,99	2,3263
1 %	0,995	2,5758

Comme $P(-t < U < t) = 1 - \alpha$, on a $P\left(-t < \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}} < t\right) = 1 - \alpha$,

$$\text{d'où } P\left(-t \frac{\sigma}{\sqrt{n}} < \bar{X} - m < t \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

$$\text{soit } P\left(m - t \frac{\sigma}{\sqrt{n}} < \bar{X} < m + t \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Ainsi l'intervalle de confiance au risque α pour la moyenne est :

$$\left[m - t \frac{\sigma}{\sqrt{n}} ; m + t \frac{\sigma}{\sqrt{n}} \right]$$

Il reste une difficulté pratique : on ne dispose pas nécessairement des paramètres (m et σ) de la loi normale représentant chacun des éléments de la population. Il est même rare de les connaître dans la mesure où justement on cherche une information sur la population.

Méthode

- Si on ne connaît pas m , on calcule la moyenne dans l'échantillon et on l'utilise dans l'intervalle de confiance.
- Si on ne connaît pas σ , on utilise la valeur s à la place de σ . Cette valeur est une estimation de l'écart-type de la population et se calcule par la formule :

$$s = \sqrt{\frac{n}{n-1} (\text{moyenne des carrés} - \text{moyenne au carré})}$$

Remarques

- Dans la formule de s , si n est très grand le coefficient, $\frac{n}{n-1}$ tend vers 1 et on retrouve un calcul classique d'écart-type.

- La construction des intervalles de confiance pour la moyenne s'est faite sur l'hypothèse d'une loi normale. Avec les approximations de lois du chapitre précédent, il est envisageable d'élargir le cadre des intervalles de confiance à d'autres lois.

7.3 ESTIMATION DE PROPORTION

a) Estimateur de proportion

On s'intéresse dans cette partie à l'estimation d'une proportion. Par exemple, on fait un sondage dans la population pour déterminer quel pourcentage va voter pour un candidat donné. On souhaite ensuite utiliser ce sondage pour conjecturer le résultat du vote.

On suppose que chaque élément de l'échantillon est représenté par une variable de Bernoulli X_i qui vaut 1 avec la probabilité p et 0 avec la probabilité $1 - p$. L'estimateur de proportion dans l'échantillon est :

$$Y = \frac{X_1 + X_2 + X_3 + \cdots + X_n}{n}$$

Remarque : cet estimateur semble être le même que le précédent, la différence est la nature des variables X_i ; elles suivent une loi normale dans le cas de l'estimateur de moyenne et une loi de Bernoulli dans le cas de l'estimateur de proportion.

Cherchons à déterminer l'espérance et la variance de Y .

$$\begin{aligned} \text{Espérance : } E(Y) &= E\left(\frac{X_1 + X_2 + X_3 + \cdots + X_n}{n}\right) \\ &= \frac{E(X_1) + E(X_2) + E(X_3) + \cdots + E(X_n)}{n} \\ &= \frac{p + p + p + \cdots + p}{n} = \frac{np}{n} = p \end{aligned}$$

Variance :

$$\begin{aligned} V(Y) &= V\left(\frac{X_1 + X_2 + X_3 + \cdots + X_n}{n}\right) \\ &= \frac{V(X_1) + V(X_2) + V(X_3) + \cdots + V(X_n)}{n^2} \\ &= \frac{p(1-p) + p(1-p) + p(1-p) + \cdots + p(1-p)}{n^2} \\ &= \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n} \end{aligned}$$

$$\text{Écart-type : } \sigma_Y = \sqrt{\frac{p(1-p)}{n}}$$

On admet que la variable Y peut-être approximée par une loi Normale de moyenne p et d'écart-type $\sqrt{\frac{p(1-p)}{n}}$

Ainsi, une estimation ponctuelle de la proportion dans la population est la proportion dans l'échantillon.

b) Intervalle de confiance

L'intervalle de confiance au risque α pour la proportion est donc

$$\left[p - t \sqrt{\frac{p(1-p)}{n}} ; p + t \sqrt{\frac{p(1-p)}{n}} \right] \text{ où } t \text{ est donné par la rela-}$$

$$\text{tion } P(U < t) = 1 - \frac{\alpha}{2}$$

Remarque : si on ne dispose pas de p dans la population, on calcule p dans l'échantillon et on l'utilise pour réaliser l'intervalle de confiance.

7.4 DIFFÉRENCES DE MOYENNES OU DE PROPORTIONS

On s'intéresse à la différence de **moyennes** entre deux populations indépendantes. Notons \bar{X}_a l'estimateur de moyenne dans l'échantillon a issu de la population A et \bar{X}_b l'estimateur de moyenne dans l'échantillon b issu de la population B . \bar{X}_a suit une loi normale de moyenne m_a et d'écart-type $\frac{\sigma_a}{\sqrt{n_a}}$ où n_a représente la taille de l'échantillon a . De

même, \bar{X}_b suit une loi normale de moyenne m_b et d'écart-type $\frac{\sigma_b}{\sqrt{n_b}}$

où n_b représente la taille de l'échantillon b . Notons $Z = \bar{X}_a - \bar{X}_b$ la différence entre les deux moyennes.

$$\text{On a } E(Z) = E(\bar{X}_a - \bar{X}_b) = E(\bar{X}_a) - E(\bar{X}_b) = m_a - m_b$$

$$\text{Et } V(Z) = V(\bar{X}_a - \bar{X}_b) = V(\bar{X}_a) + V(\bar{X}_b) = \frac{\sigma_a^2}{n_a} + \frac{\sigma_b^2}{n_b}$$

$$\text{Ainsi } \sigma_Z = \sqrt{\frac{\sigma_a^2}{n_a} + \frac{\sigma_b^2}{n_b}}$$

Ainsi l'intervalle de confiance au risque α pour Z est :

$$\left[m_a - m_b - t \sqrt{\frac{\sigma_a^2}{n_a} + \frac{\sigma_b^2}{n_b}} ; m_a - m_b + t \sqrt{\frac{\sigma_a^2}{n_a} + \frac{\sigma_b^2}{n_b}} \right]$$

Par analogie, l'intervalle de confiance pour la différence de proportion entre deux populations est :

$$\left[p_a - p_b - t \sqrt{\frac{p_a(1-p_a)}{n_a} + \frac{p_b(1-p_b)}{n_b}} ; \right. \\ \left. p_a - p_b + t \sqrt{\frac{p_a(1-p_a)}{n_a} + \frac{p_b(1-p_b)}{n_b}} \right]$$

7.5 TESTS D'HYPOTHÈSES

Les estimateurs peuvent être utilisés dans la construction de tests d'hypothèses. On prend alors le problème à l'inverse des intervalles de confiance :

1. On suppose que l'on connaît la population dans son ensemble selon un certain critère.
2. On dispose d'une nouvelle série d'informations (un échantillon) qui semble différer de la loi initiale suivie par la population.
3. On doit conclure à l'aide d'un test d'hypothèse si ce changement est le fait du hasard de l'échantillonnage ou le fait d'un changement réel des paramètres de la loi dans la population.

a) Test unilatéral

Exemple. Le nombre d'articles vendus par jour pendant l'année dans un magasin suit une loi normale de moyenne 300 et d'écart-type 20. Pendant les soldes qui durent 15 jours, le nombre moyen d'articles vendus par jour est de 315. Les soldes induisent-elles une amélioration significative des ventes ?

Cet exemple va servir à mettre en place une méthode de formulation d'un test d'hypothèse.

Notons m la moyenne de la loi normale régissant le nombre d'articles vendus pendant les soldes.

Remarque : $m \neq 315$ car 315 n'est que la moyenne observée et pas nécessairement la moyenne de la loi.

Deux hypothèses alternatives sont ici en concurrence :

- l'hypothèse nulle : H_0 « les soldes n'amènent aucun changement : $m = 300$ » ;
- l'hypothèse alternative : H_1 « Les soldes amènent un progrès dans les ventes : $m > 300$ ».

Il s'agit d'arbitrer entre ces deux hypothèses. Si l'hypothèse H_0 est vraie, quelle est la probabilité que la moyenne observée soit de 315 ou plus ? Notons \bar{X} l'estimateur de moyenne dans l'échantillon ; il suit sous H_0 une loi normale de moyenne 300 et d'écart-type $\frac{20}{\sqrt{15}} = 5,16398$

Ainsi on a :

$$\begin{aligned} P(\bar{X} > 315) &= P\left(U > \frac{315 - 300}{5,16398}\right) = P(U > 2,90) \\ &= 1 - P(U < 2,90) = 1 - 0,9981 = 0,19\% \end{aligned}$$

Cette probabilité est la probabilité de crédibilité de l'hypothèse H_0 . Ici cette probabilité est très faible et l'hypothèse H_0 peut-être rejetée sans grand risque.

La probabilité 0,19 % est la **probabilité critique** de l'hypothèse H_0 .

On peut prendre le problème à l'inverse et fixer un **seuil d'acceptation** de l'hypothèse H_0 . Ce seuil peut-être 1 %, 2 %, 5 % ou 10 %. Par exemple, si on choisit un seuil d'acceptation de 5 %, la valeur critique de la moyenne observée vérifie :

$$P(\bar{X} > X_c) = 0,05$$

$$\text{Soit } P(\bar{X} < X_c) = 0,95 \text{ et } P\left(U < \frac{X_c - 300}{5,16398}\right) = 0,95$$

À l'aide de la table inverse de loi normale, on trouve $\frac{X_c - 300}{5,16398} = 1,6449$

D'où $X_c = 308,49$

Conclusion : si la moyenne observée est inférieure à 308,49, l'hypothèse nulle est conservée ; si la moyenne est supérieure à 308,49, l'hypothèse nulle est rejetée.

Méthode

1. On formule deux hypothèses : une hypothèse nulle de non changement par rapport à la situation initiale et une hypothèse alternative d'augmentation (ou de diminution de la moyenne).

2. Si on dispose d'une observation, on détermine la probabilité critique pour laquelle cette observation est acceptable dans le cadre de l'hypothèse nulle. Plus cette probabilité critique est faible, moins l'hypothèse nulle est crédible.
3. On fixe une probabilité critique en dessous de laquelle l'hypothèse nulle sera rejetée et on calcule la valeur critique de la moyenne observée. Par comparaison entre cette valeur critique et notre observation, on rejette ou on accepte l'hypothèse nulle.

Remarques

- Dans l'exemple, l'alternative à l'hypothèse nulle est l'augmentation de la moyenne, on peut de façon symétrique envisager que le changement soit une baisse de la moyenne.
- Ce qui vient d'être présenté pour une variation de moyenne est transposable à une variation de proportion ou de différences de moyennes ou de proportions.

b) Test bilatéral

À l'inverse du test unilatéral, l'alternative ne se trouve plus d'un seul côté mais des deux cotés.

Exemple. Une machine à boisson remplit les verres selon une loi Normale de moyenne 33 cl et d'écart-type 5 cl. Sur les dix derniers verres, on constate une moyenne de contenance des verres de 30 cl. La machine est-elle déréglée ?

Le dérèglement de la machine peut-être envisagé de deux façons, soit uniquement par le fait que les verres ne semblent pas assez remplis, soit par le fait que les verres soient trop remplis et débordent. On va prendre en compte les deux possibilités ici. On formule ainsi deux hypothèses en notant m le paramètre de la loi normale représentant la moyenne des dix derniers verres remplis :

- l'hypothèse nulle H_0 est « $m = 33$ » ;
- l'hypothèse alternative H_1 est « $m > 33$ ou $m < 33$ ».

Sous l'hypothèse H_0 , les paramètres de la loi de \bar{X} sont une moyenne de 33 et un écart-type de $\frac{5}{\sqrt{10}} = 1,58114$. Comme le test est bilatéral, le rejet de l'hypothèse nulle se fait si il y a augmentation ou diminution de la moyenne. On ne cherche plus seulement la probabilité $P(\bar{X} < 30)$ mais aussi la probabilité symétrique par rapport à la moyenne $P(\bar{X} > 36)$.

Ainsi la **probabilité critique** est $P(\bar{X} < 30) + P(\bar{X} > 36)$, comme la loi normale est symétrique, la probabilité critique est

$$\begin{aligned} 2P(\bar{X} < 30) &= 2P\left(U < \frac{30 - 33}{1,58114}\right) = 2P(U < -1,90) \\ &= 2P(U > 1,9) = 2(1 - P(U < 1,9)) \\ &= 2(1 - 0,97128) = 5,74\% \end{aligned}$$

La conclusion dépend du **seuil d'acceptation** que l'on choisit : si on choisit un seuil d'acceptation de 5 %, l'hypothèse nulle est acceptable car la probabilité critique est supérieure à 5 %. Par contre, si on choisit un seuil d'acceptation à 10 %, l'hypothèse nulle n'est pas acceptable.

On peut chercher un intervalle de valeurs critiques pour lequel l'hypothèse nulle est acceptable à 5 %. Il s'agit de **l'intervalle de confiance** à 5 % pour la moyenne :

$$[33 - 1,96 \times 1,58114; 33 + 1,96 \times 1,58114] = [29,9; 36,099]$$

Ainsi l'hypothèse nulle est acceptable au seuil 5 % si l'observation est comprise entre 29,9 cl et 36,1 cl ce qui est le cas ici.

Remarque : les tests bilatéraux s'effectuent aussi pour des proportions ou de différences de moyennes ou de proportions.



POINTS CLEFS

► Intervalles de confiance et estimateurs

Estimateurs	Loi	Estimation ponctuelle	Intervalle de confiance
Moyenne	Normale $(m; \frac{\sigma}{\sqrt{n}})$	m	$[m - t \frac{\sigma}{\sqrt{n}}; m + t \frac{\sigma}{\sqrt{n}}]$
Proportion	Normale $(p; \sqrt{\frac{p(1-p)}{n}})$	p	$[\rho - t \sqrt{\frac{p(1-p)}{n}}; \rho + t \sqrt{\frac{p(1-p)}{n}}]$
Différence de moyenne	Normale $(m_a - m_b; \sqrt{\frac{\sigma_a^2}{n_a} + \frac{\sigma_b^2}{n_b}})$	$m_a - m_b$	$[m_a - m_b - t \sqrt{\frac{\sigma_a^2}{n_a} + \frac{\sigma_b^2}{n_b}}; m_a - m_b + t \sqrt{\frac{\sigma_a^2}{n_a} + \frac{\sigma_b^2}{n_b}}]$
Différence de proportion	Normale $(p_a - p_b; \sqrt{\frac{p_a(1-p_a)}{n_a} + \frac{p_b(1-p_b)}{n_b}})$	$p_a - p_b$	$[\rho_a - \rho_b - t \sqrt{\frac{p_a(1-p_a)}{n_a} + \frac{p_b(1-p_b)}{n_b}}; \rho_a - \rho_b + t \sqrt{\frac{p_a(1-p_a)}{n_a} + \frac{p_b(1-p_b)}{n_b}}]$

- ▶ t est un coefficient qui dépend du risque α . On trouve t dans la table de la loi normale centrée réduite inverse par la relation :

$$P(U < t) = 1 - \frac{\alpha}{2}$$

- ▶ Si on dispose pas de m ou de p , on choisit ceux de l'échantillon ; si on ne dispose pas de σ , on calcule le coefficient :

$$s = \sqrt{\frac{n}{n-1} (\text{moyenne des carrés} - \text{moyenne au carré})}$$

- ▶ **Test d'hypothèse**

- On pose une hypothèse nulle de non changement par rapport à une situation initiale.
- On pose une hypothèse alternative correspondant à notre observation. Elle peut être unilatérale ou bilatérale.
- ▶ On teste l'hypothèse nulle par calcul d'une probabilité critique ou d'un seuil d'acceptation de l'hypothèse nulle.
- ▶ La **probabilité critique** est la probabilité d'observée les conditions de l'hypothèse alternative sous l'hypothèse nulle.
- ▶ Le **seuil d'acceptation** est la limite au dessus (ou en dessous) de laquelle on doit rejeter l'hypothèse nulle avec un risque fixé.

EXERCICES

7.1 Recettes

Une usine produit deux objets A et B. La quantité d'objets A produite en une journée suit une loi normale de moyenne 2 000 et d'écart-type 500. La quantité d'objet B produite en une journée suit une loi normale de moyenne 1 000 et d'écart-type 300. Un objet A est vendu 30 euros et un objet B est vendu 50 euros. On suppose que tous les objets produits sont vendus.

- Donner la loi de R représentant la recette réalisée en une journée. Donner sa moyenne et son écart-type.
- Les dépenses liées à la production engendrent un coût fixe de 80 000 euros quelle que soit la production. Déterminer la probabilité que l'usine perde de l'argent en une journée.

c) On observe les recettes de cette usine sur un mois de 30 jours. Donner un intervalle de confiance à 5 % de la recette moyenne sur ces 30 jours.

d) Expliquer comment évoluerait cet intervalle de confiance si le nombre de jours d'observation augmentait.

7.2 Tueur en liberté

Un tueur est recherché par la police, il a été signalé 23 fois dans un village à 6km du lieu du crime, 47 fois dans un supermarché à 25 km du lieu du crime et 3 fois dans une autre région à 150 km du lieu du crime.

À quelle distance du lieu du crime la police doit-elle effectuer ses recherches pour avoir 95 % de chances de trouver le tueur ?

7.3 Référendum

On effectue un sondage sur 100 personnes avant un référendum, 51 personnes affirment qu'elles vont voter oui.

a) Donner un intervalle de confiance au risque 5 % de la proportion de personnes qui vont voter oui.

b) Peut-on affirmer que le oui va l'emporter ?

c) Pour quelle taille de population sondée, peut-on affirmer à 95 % que le oui va l'emporter sachant que la proportion de « votant oui » demeure constante dans l'échantillon ?

7.4 Contrôle de fabrication

On effectue un contrôle de fabrication sur des pièces dont une proportion p est défectueuse. On contrôle un lot de 400 pièces et on trouve 30 pièces défectueuses. Donner un intervalle de confiance pour l'estimation de p au niveau 95 %, puis 99 %.

7.5 Électroménager

Un vendeur d'appareils électroménagers assure un remboursement de l'appareil s'il tombe en panne avant 5 ans. Sur un échantillon de 120 appareils, on compte une durée de vie moyenne de 5 ans et demi avec un écart-type de 2 ans. Quel est le risque pour le vendeur de voir ses appareils tomber en panne en moyenne avant 5 ans ?

7.6 La crise

On compare les baisses du cours des actions réalisées au moment de la crise grecque, le vendredi 14 avril 2010, par des entreprises américaines du Dow Jones 30 et des entreprises françaises du CAC40.

Dow Jones 30	Variations	CAC40	Variations
Caterpillar	- 3,03 %	Alcatel Lucent	- 5,55 %
Coca-Cola	- 0,28 %	Danone	- 3,15 %
Mc Donald's	- 1,28 %	Bouygues	- 4,04 %
Bank of America	- 3,14 %	EDF	- 4,91 %
Hewlett Packard	- 2,65 %	Michelin	- 3,24 %
Microsoft	- 1,06 %	Sanofi-Aventis	- 2,44 %

- Donner l'intervalle de confiance au risque 5 % pour la différence des variations relatives moyennes des deux populations.
- Dans quelle mesure peut-on affirmer que le marché français souffre davantage de la crise grecque que les États-Unis ?
- Formulez un test d'hypothèses unilatéral pour calculer la valeur critique d'écart au risque 1 %.
- Quelles limites voyez-vous à ce modèle ?

7.7 Tests de qualité

Une entreprise effectue des tests de qualité sur les productions de deux de ses usines ; sur un échantillon de 400 produits de l'usine A, 2 % sont défectueux et sur un échantillon de 350 produits de l'usine B 1,5 % sont défectueux.

- Quel est l'intervalle de confiance au risque 5 % pour la différence de proportions de pièces défectueuses entre les deux usines ?
- Que peut-t-on conclure ?

7.8 Sondage

Lors des dernières élections, le parti A a réalisé 52 % des suffrages. On réalise un sondage sur 1 000 personnes et 49 % de ces personnes affirment vouloir voter pour ce parti. On s'intéresse à cette baisse de pourcentage

- Poser les hypothèses d'un test unilatéral.
- Déterminer la probabilité critique liée à l'observation.
- Déterminer la valeur critique de proportion au risque 5 %.
- Conclure.

7.9 Chiffre d'affaires

Le chiffre d'affaires journalier d'un magasin suit une loi normale de moyenne 1 000 euros et d'écart-type 500 euros. Le magasin change la présentation de sa vitrine et s'interroge sur l'effet de ce changement sur le chiffre d'affaire des 20 prochains jours.

a) Formulez un test d'hypothèse bilatéral sur la moyenne du chiffre d'affaires journalier pour déterminer un intervalle critique en dehors duquel il est possible d'affirmer au risque 1 % que le changement de présentation a induit une modification du chiffre d'affaires.

b) On constate un chiffre d'affaires moyen de 1 200 euros sur ces 20 jours. Que conclure ?

SOLUTIONS

Exercice 7.1

a) Notons X le nombre d'objets A vendus et Y le nombre d'objets B vendus. La recette est $R = 30X + 50Y$.

On a :

$$\begin{aligned} E(R) &= E(30X + 50Y) = 30E(X) + 50E(Y) \\ &= 30 \times 2\,000 + 50 \times 1\,000 = 110\,000 \text{ euros} \end{aligned}$$

$$\begin{aligned} V(R) &= V(30X + 50Y) = 30^2V(X) + 50^2V(Y) \\ &= 30^2 \times 500^2 + 50^2 \times 300^2 = 450\,000\,000 \end{aligned}$$

Ainsi l'écart-type de R est $\sigma_R = \sqrt{450\,000\,000} = 21\,213,2$ euros.

R suit donc une loi normale de moyenne 110 000 et d'écart-type 21 213,2.

b) On cherche la probabilité :

$$\begin{aligned} P(R < 80\,000) &= P\left(U < \frac{80\,000 - 110\,000}{21\,213,2}\right) = P(U < -1,41) \\ &= P(U > 1,41) = 1 - P(U < 1,41) = 1 - 0,92073 \\ &= 7,927\% \end{aligned}$$

c) L'intervalle de confiance est :

$$\begin{aligned} &\left[110\,000 - 1,96 \frac{21\,213,2}{\sqrt{30}} ; 110\,000 + 1,96 \frac{21\,213,2}{\sqrt{30}} \right] \\ &= [102\,408,95 ; 117\,591,05] \end{aligned}$$

d) Si le nombre de jours d'observations augmente, la valeur de n augmente et par conséquent la largeur de l'intervalle diminue. Si on dispose de plus d'observations, l'estimation de la moyenne sera plus précise.

Exercice 7.2

Il s'agit de faire un intervalle de confiance à 95 % pour la distance moyenne à laquelle se trouve le tueur du lieu du crime. On a $t = 1,96$, la moyenne est $m = \frac{23 \times 6 + 47 \times 25 + 3 \times 150}{23 + 47 + 3} = 24,15$ km, le coefficient

$$s = \sqrt{\frac{23+47+3}{23+47+3-1} \left(\frac{23 \times 6^2 + 47 \times 25^2 + 3 \times 150^2}{23+47+3} - 24,15^2 \right)} = 27,67 \text{ km et}$$

$$n = 23 + 47 + 3 = 73$$

Ainsi l'intervalle de confiance est :

$$\left[24,15 - 1,96 \frac{27,67}{\sqrt{73}} ; 24,15 + 1,96 \frac{27,67}{\sqrt{73}} \right] = [17,8 ; 30,5]$$

Les recherches vont pouvoir se concentrer dans un rayon d'au plus 30,5 km et d'au moins 17,8 km du lieu du crime.

Exercice 7.3

a) On a $t = 1,96$, $n = 100$ et $p = 0,51$ ainsi l'intervalle de confiance est :

$$\left[0,51 - 1,96 \sqrt{\frac{0,51(1 - 0,51)}{100}} ; 0,51 + 1,96 \sqrt{\frac{0,51(1 - 0,51)}{100}} \right]$$

$$= [41,2 \% ; 60,8 \%]$$

b) On peut affirmer au risque 5 % que le résultat du référendum sera entre 41,2 % et 60,8 %, or pour l'emporter le oui doit atteindre 50 % ou plus. On ne peut donc rien affirmer quant au résultat du vote.

c) Pour être assuré au risque 5 % que le oui l'emporte, il faudrait que la borne inférieure de l'intervalle de confiance soit supérieure à 50 %. On suppose que le pourcentage de « votants oui » demeure constant à 51 %.

$$\text{Ainsi } 0,51 - 1,96 \sqrt{\frac{(0,51(1 - 0,51))}{n}} > 0,5$$

$$\text{Soit } 0,51 - 0,5 > 1,96 \sqrt{\frac{(0,51(1 - 0,51))}{n}}$$

$$\text{D'où } \frac{0,01}{1,96} > \sqrt{\frac{(0,51(1 - 0,51))}{n}}$$

$$\text{Soit } \left(\frac{0,01}{1,96} \right)^2 > \frac{(0,51(1 - 0,51))}{n}$$

$$\text{Ainsi } n > \frac{0,51(1 - 0,51)}{\left(\frac{0,01}{1,96}\right)^2}$$

Soit $n > 9\,600,16$. Il est donc nécessaire d'interroger au moins 9 601 personnes pour s'assurer au risque 5 % du résultat du vote.

Exercice 7.4

Il s'agit d'une application directe des intervalles de confiance pour une proportion. La proportion dans l'échantillon est de $\frac{30}{400} = 0,075$

– Au niveau 95 %, on a $t = 1,96$ ainsi :

$$\begin{aligned} & \left[p - t\sqrt{\frac{p(1-p)}{n}} ; p + t\sqrt{\frac{p(1-p)}{n}} \right] \\ & = \left[0,075 - 1,96\sqrt{\frac{0,075(1-0,075)}{400}} ; 0,075 + 1,96\sqrt{\frac{0,075(1-0,075)}{400}} \right] \\ & = [0,049 ; 0,101] \end{aligned}$$

– Au niveau 99 %, on a $t = 2,5758$ ainsi :

$$\begin{aligned} & \left[p - t\sqrt{\frac{p(1-p)}{n}} ; p + t\sqrt{\frac{p(1-p)}{n}} \right] \\ & = \left[0,075 - 2,5758\sqrt{\frac{0,075(1-0,075)}{400}} ; 0,075 + 2,5758\sqrt{\frac{0,075(1-0,075)}{400}} \right] \\ & = [0,041 ; 0,108] \end{aligned}$$

Exercice 7.5

On cherche en réalité dans cet exercice un intervalle de confiance centré sur 5,5 ans et dont la borne inférieure soit de 2 ans. On dispose de l'écart-type et de la taille de l'échantillon, reste à déterminer la valeur de t liée au risque pris par le vendeur :

$$5,5 - t \frac{2}{\sqrt{120}} = 5$$

$$\text{ainsi } 5,5 - 5 = t \frac{2}{\sqrt{120}} \text{ soit } t = 0,5 \frac{\sqrt{120}}{2} = 2,74$$

On rappelle que t est donné par la relation : $P(U < t) = 1 - \frac{\alpha}{2}$

D'après la table de la loi normale centrée réduite, on a :

$$P(U < 2,74) = 0,99693$$

$$\text{Ainsi } 1 - \frac{\alpha}{2} = 0,99693 \text{ d'où } \frac{\alpha}{2} = 0,00307$$

Le risque de voir ses appareils tomber en panne en moyenne avant 5 ans est de 0,307 %, ce qui est très faible.

Exercice 7.6

a) Déterminons dans un premier temps la variation moyenne des entreprises américaines et la variation moyenne des entreprises françaises, ainsi que le coefficient s de chacune.

Pour les entreprises américaines :

$$m_a = \frac{-3,03 - 0,28 - 1,28 - 3,14 - 2,65 - 1,06}{6} = -1,91 \%$$

$$s_a = \sqrt{\frac{6}{5} \left(\frac{3,03^2 + 0,28^2 + 1,28^2 + 3,14^2 + 2,65^2 + 1,06^2}{6} - (-1,91)^2 \right)} = 1,19 \%$$

Pour les entreprises françaises :

$$m_f = \frac{-5,55 - 3,15 - 4,04 - 4,91 - 3,24 - 2,44}{6} = -3,89 \%$$

$$s_f = \sqrt{\frac{6}{5} \left(\frac{5,55^2 + 3,15^2 + 4,04^2 + 4,91^2 + 3,24^2 + 2,44^2}{6} - (-3,89)^2 \right)} = 1,17 \%$$

L'intervalle de confiance au risque 5 % pour l'écart de moyenne est donné par la formule :

$$\begin{aligned} & \left[m_a - m_b - t \sqrt{\frac{\sigma_a^2}{n_a} + \frac{\sigma_b^2}{n_b}}; m_a - m_b + t \sqrt{\frac{\sigma_a^2}{n_a} + \frac{\sigma_b^2}{n_b}} \right] \\ &= \left[-1,91 - (-3,89) - 1,96 \sqrt{\frac{1,19^2}{6} + \frac{1,17^2}{6}}; \right. \\ & \quad \left. -1,91 - (-3,89) + 1,96 \sqrt{\frac{1,19^2}{6} + \frac{1,17^2}{6}} \right] \end{aligned}$$

$$= [0,64 \% ; 3,32 \%]$$

b) L'intervalle de confiance est positif, ainsi au risque d'erreur 5 % on peut affirmer que l'écart entre les baisses des entreprises américaines et les baisses des entreprises françaises est positif. On peut donc conclure au vue de cet intervalle de confiance que les entreprises françaises ont davantage souffert de la crise grecque.

c) Notons m_a et m_f les paramètres de moyenne des entreprises américaines et françaises.

- L'hypothèse nulle H_0 est « $m_a = m_f$ ».
- L'hypothèse alternative H_1 est « $m_a > m_f$ ».

On s'intéresse à la loi de $\bar{X}_a - \bar{X}_f$. On cherche un écart critique E_c pour lequel $P(\bar{X}_a - \bar{X}_f > E_c) = 0,01$. Si on suppose l'hypothèse H_0 vrai, $\bar{X}_a - \bar{X}_f$ suit une loi normale de moyenne 0 et d'écart-type

$$\sqrt{\frac{1,19^2}{6} + \frac{1,17^2}{6}} = 0,6813. \text{ On peut ainsi « centrer réduire » la varia-}$$

$$\text{ble : } P\left(U > \frac{E_c - 0}{0,6813}\right) = 0,01, \text{ ainsi } P\left(U < \frac{E_c}{0,6813}\right) = 0,99.$$

À l'aide de la table de la loi normale inverse, on trouve :

$$\frac{E_c}{0,6813} = 2,3263$$

Ainsi $E_c = 1,5849 \%$. L'écart moyen constaté est de $-1,91 - (-3,89) = 1,98 \%$. Cet écart est supérieur à l'écart critique, au risque 1 % l'hypothèse nulle doit être rejetée.

d) Plusieurs limites se présentent à notre modèle, tout d'abord la taille de l'échantillon qui est très petit (6 entreprises américaines et 6 entreprises françaises), le choix des entreprises dans le Dow Jones ou le CAC 40 indique des entreprises de fortes capitalisations et donc non nécessairement représentatives de l'ensemble de marché national. Le calcul de la moyenne a donné un poids égal à chaque entreprises ; or, elles sont de tailles différentes, il aurait fallu pondérer soit par la taille soit par le volume de titres échangés. La construction d'un intervalle de confiance suppose des variables qui suivent une loi normale, rien n'assure que c'est le cas ici.

Exercice 7.7

a) Il s'agit de l'intervalle de confiance d'une différence de deux proportions. Le risque est 5 % ainsi $t = 1,96$, on a $n_a = 400$, $p_a = 2 \%$, $n_b = 350$ et $p_b = 1,5 \%$. L'intervalle de confiance est donné par la formule :

$$\begin{aligned} & \left[p_a - p_b - t \sqrt{\frac{p_a(1-p_a)}{n_a} + \frac{p_b(1-p_b)}{n_b}} ; p_a - p_b + t \sqrt{\frac{p_a(1-p_a)}{n_a} + \frac{p_b(1-p_b)}{n_b}} \right] \\ & = \left[0,02 - 0,015 - 1,96 \sqrt{\frac{0,02(1-0,02)}{400} + \frac{0,015(1-0,015)}{350}} ; \right. \\ & \qquad \qquad \qquad \left. 0,02 - 0,015 + 1,96 \sqrt{\frac{0,02(1-0,02)}{400} + \frac{0,015(1-0,015)}{350}} \right] \\ & = [-1,37 \% ; 2,37 \%] \end{aligned}$$

b) Au risque 5 %, on ne peut rien conclure car l'intervalle de confiance contient des valeurs positives et des valeurs négatives.

Exercice 7.8

a) On s'intéresse à la proportion dans le sondage. L'hypothèse nulle est $p = 52\%$. L'hypothèse alternative est $p < 52\%$.

b) Sous l'hypothèse H_0 , la proportion Y dans l'échantillon suit une loi normale de moyenne 52% et d'écart-type $\sqrt{\frac{0,52(1-0,52)}{1\,000}} = 1,5799\%$. La probabilité critique est donc :

$$\begin{aligned} P(Y < 0,49) &= P\left(U < \frac{0,49-0,52}{0,015799}\right) = P(U < -1,90) = P(U > 1,9) \\ &= 1 - P(U < 1,9) = 1 - 0,97128 = 2,872\% \end{aligned}$$

c) On cherche une valeur critique pour laquelle $P(Y < Y_c) = 0,05$ soit $P\left(U < \frac{Y_c - 0,52}{0,015799}\right) = 0,05$

Ainsi, d'après la table de la loi normale inverse :

$$\frac{Y_c - 0,52}{0,015799} = -1,6449 \text{ soit } Y_c = 49,4\%$$

d) La question b) indique que l'on peut donner une crédibilité de $2,87\%$ à l'hypothèse nulle, ce qui est assez faible. La question c) indique que l'on doit rejeter l'hypothèse nulle au risque 5% si la proportion observée est inférieure à $49,4\%$, ce qui est aussi le cas ici. En conclusion, le parti A subit une baisse réelle de popularité dans l'opinion.

Exercice 7.9

a) Notons m le paramètre de la moyenne de l'estimateur de la moyenne du chiffre d'affaires des 20 jours. On note H_0 l'hypothèse nulle, on a H_0 : « $m = 1\,000$ » ; et l'hypothèse alternative H_1 : « $m \neq 1\,000$ ». L'intervalle critique au seuil 1% n'est autre que l'intervalle de confiance du chiffre d'affaire moyen qui suit sous l'hypothèse H_0 une loi normale de moyenne $1\,000$ et d'écart-type $\frac{1\,000}{\sqrt{20}} = 223,61$. L'intervalle de confiance est :

$$\begin{aligned} &[1\,000 - 2,5758 \times 223,61 ; 1\,000 + 2,5759 \times 223,61] \\ &= [424,03 ; 1\,575,99] \end{aligned}$$

b) Un chiffre d'affaires moyen de $1\,200$ est dans l'intervalle critique, il est d'ailleurs assez loin des extrémités, on peut donc conclure que l'hypothèse H_0 est acceptable.

Test du χ^2

OBJECTIFS

- ▶ Connaître la formule et la méthode du χ^2 (khi-deux)
- ▶ Savoir calculer un nombre de degrés de libertés
- ▶ Lire la table de la loi du χ^2
- ▶ Utiliser le χ^2 pour conclure quant à la validité d'une modélisation
- ▶ Utiliser le χ^2 pour montrer que deux paramètres sont indépendants

PLAN

- 8.1 Modélisation par une loi
- 8.2 Test d'indépendance

L'observation d'un phénomène conduit à une recherche de modèle. On a vu qu'une loi dont l'espérance est proche de la variance invite à une modélisation par une loi de Poisson, une variable continue dont la densité ressemble à une cloche invite à une modélisation par une loi normale.

Le problème se pose ainsi :

1. On dispose d'une série d'observations.
2. On propose un modèle proche de notre observation.
3. On s'interroge sur la légitimité de notre modèle.

Le test du χ^2 va répondre au dernier point, il va permettre de conclure avec un risque donné si notre modèle est légitime ou non.

8.1 MODÉLISATION PAR UNE LOI

Reprenons l'exemple de l'exercice du chapitre 6 pour mettre en place la méthode.

Exemple. On réalise une enquête sur 200 personnes qui indiquent le nombre d'ordinateurs chez eux :

Tableau 8-1 Observations

Nombre d'ordinateurs	0	1	2	3	4	5	6 et plus
Nombre de personnes	60	73	43	17	5	2	0

On s'interroge sur la possibilité de modéliser cette observation par une loi de Poisson de paramètre 1,2 ; en effet, on constate que la moyenne est égale à la variance de cette série. Cela ne suffit pas à conclure quant à la légitimité de notre modèle.

Une idée simple est de se demander quelle aurait été la distribution si elle suivait parfaitement une loi de Poisson et de mesurer les écarts entre l'observation et le modèle théorique.

La loi de Poisson est donnée par la formule $P(X = k) = e^{-1,2} \times \frac{1,2^k}{k!}$

On trouve ainsi le nombre de personnes théoriquement dans chaque cas en multipliant cette probabilité par 200.

Voici la distribution observée :

Tableau 8-2 Effectifs calculés

Nombre d'ordinateurs	0	1	2	3	4 et plus
Nombre de personnes	60	73	43	17	7
Distribution théorique	60,25	72,28	43,38	17,34	6,76

Remarques.

La dernière case du modèle a été complétée de manière à ce que la somme des cases soit égale à 200.

On regroupe les dernières cases de manière à avoir des effectifs au moins égaux à 5.

On observe des écarts qui semblent très faibles entre la théorie et l'observation. Mais est-ce suffisant pour conclure ? Non : il est nécessaire de mettre en place un test qui légitime avec un certain risque notre modèle.

Pour cela, on introduit une variable qui se nomme χ^2 et se calcule ainsi :

$$\chi^2 = \sum \frac{(\text{Valeurs observées} - \text{Valeurs calculées})^2}{\text{Valeurs calculées}}$$

Cette grandeur mesure des « **écarts relatifs** » entre l'observation et le modèle.

Dans notre exemple, on trouve :

$$\begin{aligned} \chi^2 = & \frac{(60 - 60,24)^2}{60,24} + \frac{(73 - 72,28)^2}{72,28} + \frac{(43 - 43,38)^2}{43,38} \\ & + \frac{(17 - 17,34)^2}{17,34} + \frac{(7 - 6,76)^2}{6,76} = 0,026644 \end{aligned}$$

Pour savoir si cette valeur est faible ou non, il faut la comparer à une valeur de référence donnée par la table du χ^2 :

Tableau 8-3 Table du χ^2

ddl	$\alpha = 5 \%$	$\alpha = 1 \%$	$\alpha = 0,1 \%$	ddl	$\alpha = 5 \%$	$\alpha = 1 \%$	$\alpha = 0,1 \%$
1	3,84	6,64	10,83	21	32,67	38,93	46,80
2	5,99	9,21	13,82	22	33,92	40,29	48,27
3	7,82	11,35	16,27	23	35,17	41,64	49,73
4	9,49	13,28	18,47	24	36,42	42,98	51,18
5	11,07	15,09	20,52	25	37,65	44,31	52,62
6	12,59	16,81	22,46	26	38,89	45,64	54,05
7	14,07	18,48	24,32	27	40,11	46,96	55,48
8	15,51	20,09	26,13	28	41,34	48,28	56,89
9	16,92	21,67	27,88	29	42,56	49,59	58,30
10	18,31	23,21	29,59	30	43,77	50,89	59,70
11	19,68	24,73	31,26	31	44,99	52,19	61,10
12	21,03	26,22	32,91	32	46,19	53,49	62,49
13	22,36	27,69	34,53	33	47,40	54,78	63,87
14	23,69	29,14	36,12	34	48,60	56,06	65,25
15	25,00	30,58	37,70	35	49,80	57,34	66,62
16	26,30	32,00	39,25	36	51,00	58,62	67,99
17	27,59	33,41	40,79	37	52,19	59,89	69,35
18	28,87	34,81	42,31	38	53,38	61,16	70,71
19	30,14	36,19	43,82	39	54,57	62,43	72,06
20	31,41	37,57	45,32	40	55,76	63,69	73,41

La première ligne de la table indique le risque de rejeter à tort le modèle alors que celui-ci est compatible avec l'observation. On envisage généralement des risques de 5 %, 1 % et 0,1 %.

La première colonne indique le nombre de **degrés de libertés** (ddl) du problème. Le nombre de degrés de libertés se calcule par la formule :

$$\text{ddl} = \text{Nombre de cases après regroupement} - 1 \\ - \text{Nombre de paramètres de la loi}$$

Le nombre de cases de cet exemple est 5 après regroupement. La loi de Poisson a un seul paramètre : λ (la loi normale en a deux : m et σ)

Ainsi on dispose de $5 - 1 - 1 = 3$ degrés de libertés.

Si on fixe un risque de 1 %, la valeur limite du χ^2 dans la table est 11,35

La règle est ensuite la suivante :

- si χ^2 observé $>$ χ^2 de la table alors on rejette la modélisation ;
- si χ^2 observé $<$ χ^2 de la table alors on légitime la modélisation.

Ici, très nettement, on a χ^2 observé $<$ χ^2 de la table ($0,02664 < 11,35$) donc le modèle est légitime.



Méthode

1. On calcule des effectifs théoriques à l'aide de la loi proposée dans le modèle et on arrondit le dernier de manière à ce que la somme soit égale à l'effectif total.
2. On regroupe les valeurs adjacentes de manière à ce que les effectifs dans les cases soit au moins de 5.
3. On calcule le c^2 .
4. On choisit un risque et on calcule le degré de liberté.
5. On compare la valeur observée à la valeur de la table de manière à conclure si, oui ou non, le modèle est légitime.

8.2 TEST D'INDÉPENDANCE

Le test du χ^2 permet aussi de vérifier l'indépendance ou non de deux facteurs.

Exemple. On s'interroge sur le lien entre le salaire et le fait d'avoir moins de 30 ans dans une entreprise. Le *Tableau 8-4*, page suivante, présente les résultats.

Tableau 8-4 Répartition des salaires

	Salaire < 1 500 euros	Salaire ≥ 1 500 euros	Total
Moins de 30 ans	68	35	103
Plus de 30 ans	70	50	120
Total	138	85	223

Le salaire est-il indépendant de la classe d'âge dans cette entreprise ?

Dans un premier temps, nous allons calculer ce qu'aurait été la répartition des employés si les deux événements étaient indépendants.

Pour cela, on utilise la formule :

$$\frac{\text{Bout de la ligne} \times \text{Bout de la colonne}}{\text{Total des cases}}$$

Par exemple, pour la première case on trouve l'effectif théorique par :

$$\frac{103 \times 138}{223} = 63,74$$

On complète ainsi le tableau :

Tableau 8-5 Répartition théorique des salaires

	Salaire < 1500 euros	Salaire > 1500 euros	Total
Moins de 30 ans	63,75	39,26	103
Plus de 30 ans	74,26	45,74	120
Total	138	85	223

On calcule ensuite le χ^2 entre les deux tableaux :

$$\chi^2 = \frac{(63,74 - 68)^2}{63,74} + \frac{(39,26 - 35)^2}{39,26} + \frac{(70 - 74,26)^2}{74,26} + \frac{(50 - 45,74)^2}{45,74}$$

$$= 1,39$$

Dans le cas du test d'indépendance, le nombre de degrés de libertés du problème est donné par la formule :

$$\text{ddl} = (\text{Nombre de lignes} - 1) \times (\text{Nombre de colonnes} - 1)$$

Remarque : on ne compte pas les lignes et les colonnes « Total » dans la formule.

Ainsi le nombre de degrés de libertés est ici : $(2 - 1) \times (2 - 1) = 1$

Au risque 1 % de rejeter l'hypothèse d'indépendance alors que celle-ci est bonne, la table nous donne un χ^2 de 6,64.

Notre observation est inférieure à cette valeur ainsi on peut conclure que la répartition des salaires est indépendante de la classe d'âge dans cette entreprise.

La méthode de comparaison est la même que dans la partie précédente :

- si χ^2 observé $>$ χ^2 de la table alors on rejette l'indépendance ;
- si χ^2 observé $<$ χ^2 de la table alors on accepte l'indépendance.



POINTS CLEFS

- Le χ^2 se calcule par la formule

$$\chi^2 = \sum \frac{(\text{Valeurs observées} - \text{Valeurs calculées})^2}{\text{Valeurs calculées}}$$

- Le nombre de **degrés de libertés** se calcule par la formule :

Pour un test de modèle,

ddl = Nombre de cases après regroupement - 1 - Nombre de paramètres de la loi

Pour un test d'indépendance,

ddl = (Nombre de lignes - 1) \times (Nombre de colonnes - 1)

- Dans la table du χ^2 , α représente le risque de rejeter à tort le modèle (ou l'indépendance) alors que celui-ci (ou celle-ci) est légitime.
- Le choix s'effectue ainsi :
 - si χ^2 observé $>$ χ^2 de la table alors on rejette le modèle ou l'indépendance ;
 - si χ^2 observé $<$ χ^2 de la table alors on accepte le modèle ou l'indépendance.

EXERCICES

8.1. Répartition des salaires

Dans une entreprise, on observe la répartition des salaires suivante :

Salaires	[800 ; 1 000[[1 000 ; 1 200[[1 200 ; 1 500[[1 500 ; 1 800[
Effectifs	10	40	100	110

Salaires	[1 800 ; 2 500[[2 500 ; 3 000[[3 000 ; 4 000[
Effectifs	30	8	2

- a) Si on pouvait modéliser la répartition des salaires par une loi normale quels en serait les paramètres ?
- b) À l'aide d'un test du χ^2 , déterminer au risque 1 % de rejeter à tort le modèle, si le modèle d'une loi normale est compatible avec l'observation.

8.2. Jeu de dé

On lance un dé plusieurs fois et on dénombre les valeurs obtenues :

Valeurs	1	2	3	4	5	6
Nombre de lancers	10	21	15	12	23	19

À l'aide d'un test du χ^2 au risque $\alpha = 5\%$, déterminer si ce dé est équilibré ou non.

8.3. Enquête de satisfaction

On a interrogé des habitants de Limoges, de Brive et de Guéret sur l'appréciation de 4 stations de radio. Le croisement de ces deux variables donne le tableau de contingence suivant :

	Limoges	Brive	Guéret	Total
Radio1	20	17	10	47
Radio2	17	16	12	45
Radio3	39	24	12	75
Radio4	12	8	13	33
Total	88	65	47	200

Déterminer si l'appréciation d'une radio dépend ou non de la localisation géographique.

SOLUTIONS

Exercice 8.1

a) On calcule la moyenne et l'écart-type de la série :

La moyenne est :

$$\frac{10 \times 900 + 40 \times 1\,100 + 100 \times 1\,350 + 110 \times 1\,650 + 30 \times 2\,150 + 8 \times 2\,750 + 2 \times 3\,500}{300} \\ = 1\,543,33$$

L'écart-type est :

$$\sqrt{\frac{10 \times 900^2 + 40 \times 1\,100^2 + 100 \times 1\,350^2 + 110 \times 1\,650^2 + 30 \times 2\,150^2 + 8 \times 2\,750^2 + 2 \times 3\,500^2}{300} - 1\,543,33^2} \\ = 397,23$$

On va essayer d'assimiler cette série à une distribution normale de moyenne 1 543,33 et d'écart-type 397,23

b) Dans un premier temps, on détermine les effectifs théoriques des différentes classes.

– Pour la classe [800 ; 1 000[:

$$\begin{aligned} P(800 < X < 1\,000) \\ &= P(X < 1\,000) - P(X < 800) \\ &= P\left(U < \frac{1\,000 - 1\,543,33}{397,23}\right) - P\left(U < \frac{800 - 1\,543,33}{397,23}\right) \\ &= P(U < -1,37) - P(U < -1,87) \\ &= 1 - P(U < 1,37) - (1 - P(U < 1,87)) \\ &= P(U < 1,87) - P(U < 1,37) = 0,96926 - 0,91466 = 0,0546 \end{aligned}$$

L'effectif théorique est donc : $300 \times 0,0546 = 16,38$

– Pour la classe [1 000 ; 1 200[:

$$\begin{aligned} P(1\,000 < X < 1\,200) &= P(U < 1,37) - P(U < 0,86) \\ &= 0,91466 - 0,80511 = 0,10955 \end{aligned}$$

L'effectif théorique est donc : $300 \times 0,10955 = 32,87$

– Pour la classe [1 200 ; 1 500[:

$$\begin{aligned} P(1\,200 < X < 1\,500) &= P(U < 0,86) - P(U < 0,11) \\ &= 0,80511 - 0,5438 = 0,2613 \end{aligned}$$

L'effectif théorique est donc : $300 \times 0,2613 = 78,39$

– Pour la classe [1 500 ; 1 800[:

$$\begin{aligned} P(1\,500 < X < 1\,800) &= P(U < 0,65) - (1 - P(U < 0,11)) \\ &= 0,74215 - 1 + 0,5438 = 0,28595 \end{aligned}$$

L'effectif théorique est donc : $300 \times 0,28595 = 85,79$

– Pour la classe [1 800 ; 2 500[:

$$P(1\,800 < X < 2\,500) = 0,99202 - 0,74215 = 0,24987$$

L'effectif théorique est donc : $300 \times 0,24987 = 74,96$

– On regroupe les deux dernières classes de manière à ce que les effectifs ne soient pas inférieur à 5. Il reste théoriquement dans cette dernière classe $300 - 74,96 - 85,79 - 78,39 - 32,87 - 16,38 = 11,61$

On résume les résultats puis on calcule le χ^2 de chaque classe :

Salaires	[800 ; 1 000[[1 000 ; 1 200[[1 200 ; 1 500[
Effectifs	10	40	100
Effectifs théoriques	16,38	32,87	78,39
χ^2	$\frac{(10-16,38)^2}{16,38} = 2,49$	1,55	5,96

Salaires	[1 500 ; 1 800[[1 800 ; 2 500[[2 500 ; 4 000[
Effectifs	110	30	10
Effectifs théoriques	85,79	74,96	11,61
χ^2	6,83	26,97	0,22

Le χ^2 observé est donc :

$$2,49 + 1,55 + 5,96 + 6,83 + 26,97 + 0,22 = 44,02$$

La loi normale a deux paramètres, le nombre de cases après regroupement est 6. Le nombre de degrés de liberté est donc $6 - 1 - 2 = 3$ degrés de liberté.

À 1 % la table indique une valeur de χ^2 de : 11,35.

On a dans cet exemple χ^2 observé $>$ χ^2 de la table, on rejette donc la modélisation de cette série par une loi normale.

Exercice 8.2

On a réalisé 100 lancers ; si le dé était parfaitement équilibré, on devrait

trouver « théoriquement » $\frac{100}{6} = 16,67$ fois chaque face du dé.

On calcule ensuite le χ^2 entre l'observation et la théorie :

Valeurs	1	2	3	4	5	6
Nombre de lancers	10	21	15	12	23	19
Théorie	16,67	16,67	16,67	16,67	16,67	16,65
χ^2	2,67	1,12	0,17	1,3	2,4	0,33

Remarque : la dernière valeur théorique est de 16,65 de manière à atteindre un total de 100.

La χ^2 observé est donc $2,67 + 1,12 + 0,17 + 1,3 + 2,4 + 0,33 = 7,99$
Le nombre de cases est 6, le nombre de paramètres est 0. Le nombre de degrés de liberté est donc 5.

À 5 %, la table indique une valeur de χ^2 de 11,07

On a χ^2 observé $<$ χ^2 de la table donc on peut admettre que le dé est bien équilibré.

Exercice 8.3

Effectifs théoriques :

	Limoges	Brive	Guéret	Total
Radio1	20,68	15,275	11,045	47
Radio2	19,8	14,625	10,575	45
Radio3	33	24,375	17,625	75
Radio4	14,52	10,725	7,755	33
Total	88	65	47	200

Calculs de χ^2 :

	Limoges	Brive	Guéret
Radio1	0,022	0,195	0,099
Radio2	0,396	0,129	0,192
Radio3	1,09	0,006	0,795
Radio4	0,437	0,692	3,547

En additionnant les valeurs du dernier tableau, on trouve un χ^2 observé de 8,6

Le nombre de lignes est 4, le nombre de colonnes est 3, le nombre de degrés de liberté est donc $(4 - 1)(3 - 1) = 6$ degrés de liberté.

À 5 %, on trouve un χ^2 de 12,59

On a χ^2 observé $<$ χ^2 de la table donc on admet l'hypothèse d'indépendance de la radio et de la localisation géographique.

ÉTUDE DE CAS

Problème de synthèse

Voici les cours de clôture de plusieurs actions pendant la semaine boursière 1 :

Tableau 9-1 Comparaison d'actions

Jours	Action A	Action B	Action C	Action D
Lundi	10,1	5,3	59,3	15,5
Mardi	9,8	5,3	56,6	18,3
Mercredi	9,7	5,4	53	20,3
Jeudi	10,2	5,3	61,3	14,3
Vendredi	10,8	5,4	64,7	11,2

1. Déterminer pour chaque titre le pourcentage d'augmentation (ou de diminution) entre le lundi et le vendredi. En déduire un pourcentage d'augmentation (ou de diminution) journalier moyen par titre.
2. Un investisseur souhaite acheter des actions d'une seule catégorie. Quel sera son choix s'il souhaite des titres peu risqués ? Même question pour des titres potentiellement très rentables ? Justifier toute affirmation par un calcul.
3. Un investisseur souhaite acheter deux catégories d'actions pour se constituer un portefeuille.
 - a) Comment va-t-il s'y prendre si il souhaite minimiser le risque ? Maximiser la rentabilité ?
 - b) Choisissez un couple de titres qui conviendra à une minimisation du risque et un couple qui conviendra à une maximisation de la rentabilité.Justifier toute affirmation par un calcul.

On s'intéresse aux évolutions du titre : action A. Voici les cotations de la semaine suivante (semaine 2) :

Tableau 9-2

Jours	Action A
Lundi	10,5
Mardi	10
Mercredi	9,9
Jeudi	10,7
Vendredi	11,4

- Représenter graphiquement le cours de l'action A les semaines 1 et 2 (on ne prendra pas en compte les samedis et dimanches).
- Déterminer la droite des moindres carrés de cette série.
- Quel est le modèle de cette série ?
- À l'aide de la méthode des coefficients saisonniers, donner une estimation des cours de l'action A pour les jours de la semaine 3.
- Un investisseur achète des actions le mercredi de la semaine 1 pour les revendre le vendredi de la semaine 3. Quel serait (en pourcentage) son gain ?
- Déterminer les moyennes mobiles de période 3 pour l'action A sur les dix jours. Déterminer les jours pour lesquels le conseil est à l'achat et les jours pour lesquels le conseil est à la vente, sachant qu'un conseil à l'achat est donné quand la courbe coupe les moyennes mobiles du bas vers le haut et qu'un conseil de vente est donné quand la courbe coupe les moyennes mobiles du haut vers le bas.

Remarque : on calculera les moyennes mobiles sur les trois derniers jours.

- Quel retracement peut-t-on envisager avec un ratio de 50 % ?
- Sur les 200 derniers jours les cours de l'action sont répartis ainsi :

Tableau 9-3

Cours	[7,5 ; 8[[8 ; 8,5[[8,5 ; 9[[9 ; 9,5[[9,5 ; 10[
Nombre de jours	4	9	20	64	66
Cours	[10 ; 10,5[[10,5 ; 11[[11 ; 11,5[[11,5 ; 12[
Nombre de jours	22	8	5	2	

Déterminer la médiane et les quartiles de cette série.

12. Quels seraient les paramètres de la loi normale représentant cette série ?
13. En admettant cette modélisation, quelle est la probabilité de voir le cours supérieur à 10 un jour donné ?
14. Donner un intervalle de confiance au risque 5 % de la moyenne des cours.
15. À l'aide d'un test d'hypothèses unilatéral déterminer si les cours de l'action A constatés les 5 jours de la semaine boursière 2 constituent ou non un progrès de l'action.
16. À l'aide d'un test du χ^2 évaluer la pertinence d'une modélisation de cette série par une loi normale.
17. Un investisseur décide d'acheter chaque jour de la première semaine pour 10 000 euros d'actions A et d'actions D dans les proportions respectives de 80 % et de 20 %. Les actions D sont dites « actions de sécurité » pour le portefeuille de cet actionnaire. Comment expliquer cette expression ?
18. Quelle est la valeur globale du portefeuille de cet investisseur ?
19. Quels sont les cours moyen des actions A et des actions D dans le portefeuille de cet actionnaire ?
20. Comparer ces valeurs aux valeurs du cours moyen des actions A et D pendant la première semaine.
21. Donner l'équation de la droite des moindres carrés de l'action A en fonction de l'action D.
22. Le cours de l'action A est de 11,4 euros le vendredi de la semaine 2. Quel serait le cours de l'action D *a priori* ?
23. Si l'investisseur vend l'ensemble des titres de son portefeuille le vendredi de la semaine 2, va-t-il réaliser un profit ou une perte ? Comment l'évaluer ?
24. Que penser d'un tel modèle ?

SOLUTIONS

1. **Pour l'action A :** le pourcentage d'augmentation est donné par $\frac{10,8}{10,1} = 1,06931$. Ainsi le cours a progressé de 6,93 %. La variation a lieu sur 5 jours mais il y a seulement 4 variations : du lundi au mardi, du mardi au mercredi, du mercredi au jeudi et du jeudi au vendredi. Le coefficient moyen de variation c est donné par :

$c^4 = 1,06931$ soit $c = 1,06931^{\frac{1}{4}} = 1,01689$. Le cours a ainsi augmenté en moyenne de 1,69 % par jour.

Pour l'action B : le pourcentage d'augmentation est donné par $\frac{5,4}{5,3} = 1,01887$. Ainsi le cours a progressé de 1,89 %.

Le coefficient moyen de variation c est donné par : $c^4 = 1,01887$ soit $c = 1,01887^{\frac{1}{4}} = 1,00468$. Le cours a ainsi augmenté en moyenne de 0,47 % par jour.

Pour l'action C : le pourcentage d'augmentation est donné par $\frac{64,7}{59,3} = 1,09106$. Ainsi le cours a progressé de 9,11 %. Le coefficient

moyen de variation c est donné par : $c^4 = 1,09106$ soit $c = 1,09106^{\frac{1}{4}} = 1,02203$. Le cours a ainsi augmenté en moyenne de 2,20 % par jour.

Pour l'action D : le pourcentage de diminution est donné par $\frac{11,2}{15,5} = 0,72258$. Ainsi le cours a baissé de 27,74 %. Le coefficient

moyen de variation c est donné par : $c^4 = 0,72258$ soit $c = 0,72258^{\frac{1}{4}} = 0,92198$. Le cours a ainsi baissé en moyenne de 7,80 % par jour.

2. On ne va pas chercher à spéculer sur la hausse ou la baisse des titres. Le risque est lié aux variations relatives des cours des titres. Il s'agit donc dans cette question de déterminer quelle action a le coefficient de variation (CV) le plus faible pour déterminer l'action la moins risquée (celle qui varie le moins).

Pour l'action A : la moyenne est

$$\frac{10,1 + 9,8 + 9,7 + 10,2 + 10,8}{5} = 10,12$$

L'écart-type est $\sqrt{\frac{10,1^2 + 9,8^2 + 9,7^2 + 10,2^2 + 10,8^2}{5} - 10,12^2} = 0,3868$

Ainsi le coefficient de variation est $CV_A = \frac{0,3868}{10,12} = 0,0382$

Pour l'action B : la moyenne est

$$\frac{5,3 + 5,3 + 5,4 + 5,3 + 5,4}{5} = 5,34$$

L'écart-type est $\sqrt{\frac{5,3^2+5,3^2+5,4^2+5,3^2+5,4^2}{5} - 5,34^2} = 0,04899$

Ainsi le coefficient de variation est $CV_B = \frac{0,04899}{5,34} = 0,0092$

Pour l'action C : la moyenne est

$$\frac{59,3+56,6+53+61,3+64,7}{5} = 58,98$$

L'écart-type est $\sqrt{\frac{59,3^2+56,6^2+53^2+61,3^2+64,7^2}{5} - 58,98^2} = 3,99$

Ainsi le coefficient de variation est $CV_C = \frac{3,99}{58,98} = 0,0677$

Pour l'action D : la moyenne est

$$\frac{15,5+18,3+20,3+14,3+11,2}{5} = 15,92$$

L'écart-type est $\sqrt{\frac{15,5^2+18,3^2+20,3^2+14,3^2+11,2^2}{5} - 15,92^2} = 3,16$

Ainsi le coefficient de variation est $CV_D = \frac{3,16}{15,92} = 0,1985$

On a $CV_D > CV_C > CV_A > CV_B$, ainsi les titres B sont les moins risqués et les titres D sont les plus risqués. Des titres plus risqués peuvent éventuellement être aussi plus rentables.

3. a) Pour minimiser le risque, on va chercher à avoir des titres qui n'évoluent pas dans le même sens. En effet, si deux titres évoluent dans le même sens, le mouvement de l'un est amplifié par l'autre donc le risque global du portefeuille augmente alors que lorsque deux titres évoluent en sens contraire la baisse d'un titre peut être compensée par la hausse de l'autre titre.
- b) Un calcul de covariance va permettre de trouver le couple de titres qui évolue « le plus en sens contraire ».

Pour le couple A-B la covariance est

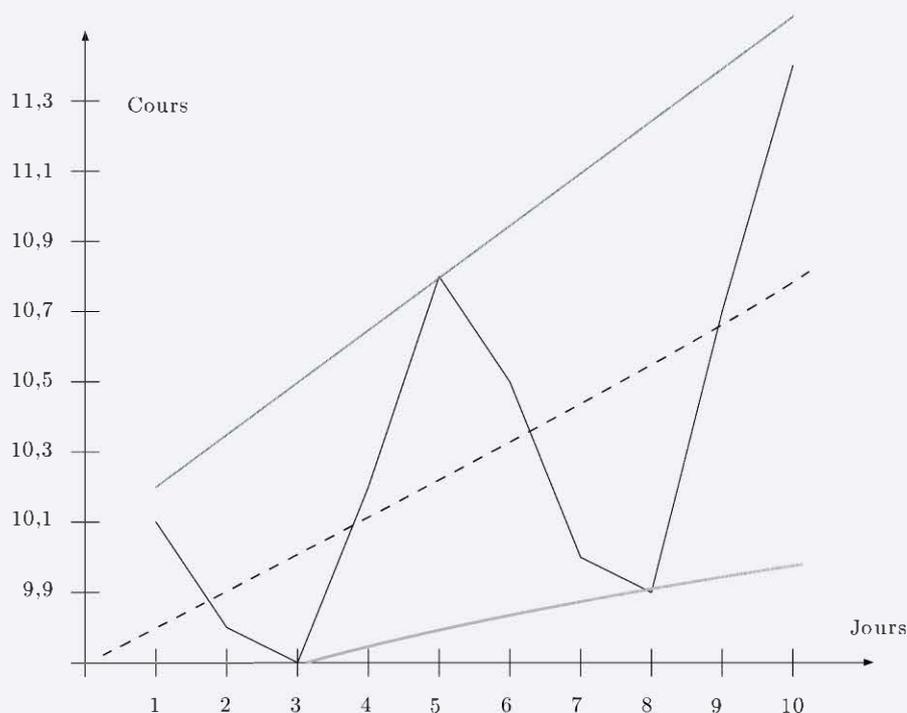
$$\frac{10,1 \times 5,3 + 9,8 \times 5,3 + 9,7 \times 5,4 + 10,2 \times 5,3 + 10,8 \times 5,4}{5} - 10,12 \times 5,34 = 0,0052$$

Résumons les covariances des autres couples dans le tableau suivant :

Couples	A-B	A-C	A-D	B-C	B-D	C-D
Covariance	0,0052	1,4682	-1,1516	-0,0052	-0,0068	-12,5474

Le couple $C-D$ est le couple qui évolue le plus significativement en sens inverse car la covariance de ce couple est négative. Un portefeuille constitué de ces deux titres est globalement peu risqué alors que chaque titre est individuellement risqué.

4. Représentation graphique de la série :



5. La droite des moindres carrés est $y = 0,1097x + 9,7067$. Elle est représentée sur le graphique et représente la tendance t de la série.
6. On trace les droites de support et de résistance de la série. On remarque que ces droites sont divergentes. Le modèle est donc géométrique.
7. Il s'agit dans un premier temps de calculer les coefficients saisonniers de chaque jour de la semaine :

x_i	y_i	t_i	$\frac{y_i}{t_i}$
1	10,1	9,82	1,03
2	9,8	9,93	0,99
3	9,7	10,04	0,97
4	10,2	10,15	1,00
5	10,8	10,26	1,05
6	10,5	10,37	1,01
7	10	10,48	0,95
8	9,9	10,58	0,94
9	10,7	10,69	1,00
10	11,4	10,80	1,06

On peut ainsi calculer les coefficients de chaque jour :

$$\text{– Lundi : } C_l = \frac{1,03 + 1,01}{2} = 1,02$$

$$\text{– Mardi : } C_m = \frac{0,99 + 0,95}{2} = 0,97$$

$$\text{– Mercredi : } C_{mer} = \frac{0,97 + 0,94}{2} = 0,955$$

$$\text{– Jeudi : } C_j = \frac{1,00 + 1,00}{2} = 1,00$$

$$\text{– Vendredi : } C_v = \frac{1,05 + 1,06}{2} = 1,055$$

Estimations pour la semaine 3 :

Jour	t_i	Coefficient	Estimations
Lundi	10,91	1,02	11,13
Mardi	11,02	0,97	10,69
Mercredi	11,13	0,955	10,63
Jeudi	11,24	1,00	11,24
Vendredi	11,35	1,055	11,97

8. Cet investisseur achète l'action à 9,7 pour la revendre à 11,97. Son gain est de $\frac{11,97}{9,7} = 1,23$; soit un gain de 23 %.

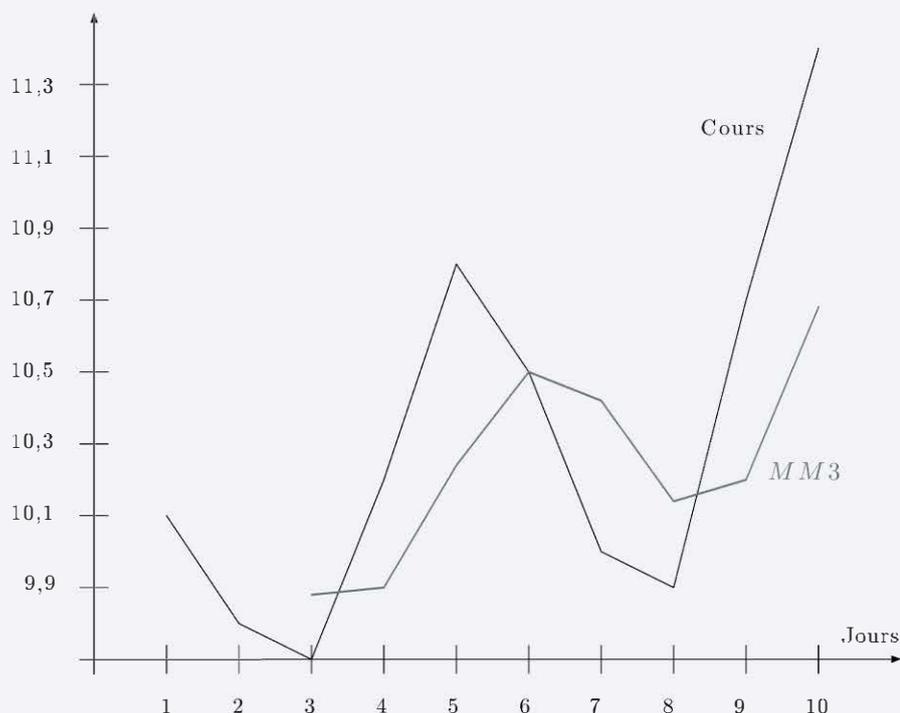
9. Calcul des moyennes mobiles :

Jours	1	2	3	4	5	6	7	8	9	10
Cours	10,1	9,8	9,7	10,2	10,8	10,5	10	9,9	10,7	11,4
MM3	-	-	9,87	9,9	10,23	10,5	10,43	10,13	10,20	10,67

On réalise les calculs des moyennes mobiles sur les trois dernières valeurs. Par exemple le premier calcul dans la troisième case est :

$$\frac{10,1 + 9,8 + 9,7}{3} = 9,87$$

Représentation graphique des moyennes mobiles :



Avec les critères de décisions de l'énoncé on conseille :

- Achat le jour 4 à 10,2 euros.
- Vente le jour 6 à 10,5 euros.
- Achat le jour 9 à 10,7 euros.

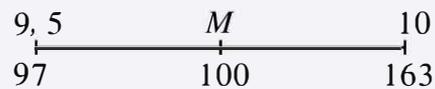
10. Entre le jour 3 et le jour 10, le cours de l'action a augmenté de 1,7 euros ($11,4 - 9,7$). Avec un ratio de 50 %, on peut envisager une baisse de $0,5 \times 1,7 = 0,85$ euros. Le cours de l'action peut ainsi descendre jusqu'à $11,4 - 0,85 = 10,55$.

11. Introduisons une ligne pour les effectifs cumulés de cette série dans le tableau :

Cours	[7,5 ; 8[[8 ; 8,5[[8,5 ; 9[[9 ; 9,5[[9,5 ; 10[
Nombre de jours	4	9	20	64	66
Effectifs cumulés	4	13	33	97	163

Cours	[10 ; 10,5[[10,5 ; 11[[11 ; 11,5[[11,5 ; 12[
Nombre de jours	22	8	5	2
Effectifs cumulés	185	193	198	200

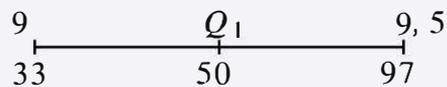
La **médiane** correspond à un effectif cumulé de 100. On a la répartition suivante :



$$\text{Ainsi : } \frac{M - 9,5}{100 - 97} = \frac{10 - 9,5}{163 - 97}$$

$$\text{d'où } M - 9,5 = 3 \times \frac{0,5}{66}, \text{ soit } M = 9,5227$$

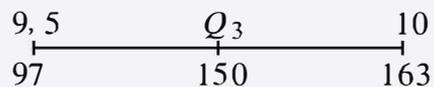
Le **premier quartile** correspond à un effectif cumulé de 50. On a la répartition suivante :



$$\text{Ainsi : } \frac{Q_1 - 9}{50 - 33} = \frac{9,5 - 9}{97 - 33}$$

$$\text{d'où } Q_1 - 9 = 17 \times \frac{0,5}{64}, \text{ soit } Q_1 = 9,1328$$

Le **troisième quartile** correspond à un effectif cumulé de 150. On a la répartition suivante :



$$\text{Ainsi : } \frac{Q_3 - 9,5}{150 - 97} = \frac{10 - 9,5}{163 - 97}$$

$$\text{d'où } Q_3 - 9,5 = 53 \times \frac{0,5}{66}, \text{ soit } Q_3 = 9,9015$$

12. Une loi normale représentant cette série aurait la même moyenne et le même écart-type que la série.

Moyenne

$$\begin{aligned} &= \frac{7,75 \times 4 + 8,25 \times 9 + 8,75 \times 20 + 9,25 \times 64 + 9,75 \times 66 + 10,25 \times 22 + 10,75 \times 8 + 11,25 \times 5 + 11,75 \times 2}{200} \\ &= 9,535. \end{aligned}$$

Écart-type

$$\begin{aligned} &= \sqrt{\frac{7,75^2 \times 4 + 8,25^2 \times 9 + 8,75^2 \times 20 + 9,25^2 \times 64 + 9,75^2 \times 66 + 10,25^2 \times 22 + 10,75^2 \times 8 + 11,25^2 \times 5 + 11,75^2 \times 2}{200} - 9,535^2} \\ &= 0,6919 \end{aligned}$$

13. On suppose dans cette modélisation que le cours X de l'action A suit une loi normale de moyenne 9,535 et d'écart-type 0,6919. La probabilité d'avoir un cours supérieur à 10 est :

$$\begin{aligned} P(X > 10) &= P\left(U > \frac{10 - 9,535}{0,6919}\right) = P(U > 0,6721) \\ &= 1 - P(U < 0,6721) \end{aligned}$$

D'après la table de la loi normale centrée réduite, on a :

$$P(U < 0,67) = 0,74857$$

Ainsi $P(X > 10) = 1 - 0,74857 = 25,14 \%$

14. Sur une période de 200 jours, l'intervalle de confiance pour la moyenne des cours est donné par $\left[m - t \frac{\sigma}{\sqrt{n}} ; m + t \frac{\sigma}{\sqrt{n}}\right]$ où $m = 9,535$, $\sigma = 0,6919$, $n = 200$ et $t = 1,96$.

L'intervalle de confiance est donc [9,439 ; 9,631].

15. Formulons le test d'hypothèses unilatéral :

- l'hypothèse nulle H_0 est « $m = 9,535$ » ;
- l'hypothèse alternative H_1 est « $m > 9,535$ ».

On suppose que l'écart-type demeure constant à 0,6919. La deuxième semaine d'observation on a une moyenne de :

$$\frac{10,5 + 10 + 9,9 + 10,7 + 11,4}{5} = 10,5$$

Quelle est la probabilité d'observer une telle moyenne sous l'hypothèse H_0 ?

On cherche la probabilité critique : $P_c = P(X > 10,5)$ sachant que X suit une loi normale de moyenne 9,535 et d'écart-type 0,6919

$$\frac{0,6919}{\sqrt{5}} = 0,3094.$$

On a donc :

$$\begin{aligned} P_c &= P\left(U > \frac{10,5 - 9,535}{0,3094}\right) = P(U > 3,12) \\ &= 1 - P(U < 3,12) = 1 - 0,99910 = 0,09 \% \end{aligned}$$

Cette probabilité étant très faible, on rejette l'hypothèse nulle au risque 0,09 %.

16. Il reste à vérifier si la loi normale était bien adaptée à cette étude.

Pour cela déterminons des effectifs théoriques des intervalles du *Tableau 10-3* en regroupant les classes d'effectifs inférieurs à 5.

– Pour la classe [7,5 ; 8,5] :

$$\begin{aligned}
 P(7,5 < X < 8,5) &= P(X < 8,5) - P(X < 7,5) \\
 &= P\left(U < \frac{8,5 - 9,535}{0,6919}\right) - P\left(U < \frac{7,5 - 9,535}{0,6919}\right) \\
 &= P(U < -1,5) - P(U < -2,94) \\
 &= 1 - P(U < 1,5) - (1 - P(U < 2,94)) \\
 &= P(U < 2,94) - P(U < 1,5) \\
 &= 0,99836 - 0,93319 = 0,06517
 \end{aligned}$$

L'effectif théorique est donc $0,06517 \times 200 = 13,03$

– Pour la classe [8,5 ; 9] :

$$\begin{aligned}
 P(8,5 < X < 9) &= P(X < 9) - P(X < 8,5) \\
 &= P\left(U < \frac{9 - 9,535}{0,6919}\right) - P(U < -1,5) \\
 &= P(U < -0,77) - P(U < -1,5) \\
 &= P(U < 1,5) - P(U < 0,77) \\
 &= 0,93319 - 0,77935 = 0,15384
 \end{aligned}$$

L'effectif théorique est donc $0,15384 \times 200 = 30,77$

– Pour la classe [9 ; 9,5] :

$$\begin{aligned}
 P(9 < X < 9,5) &= P(X < 9,5) - P(X < 9) \\
 &= P\left(U < \frac{9,5 - 9,535}{0,6919}\right) - P(U < -0,77) \\
 &= P(U < -0,05) - P(U < -0,77) \\
 &= P(U < 0,77) - P(U < 0,05) \\
 &= 0,77935 - 0,51994 = 0,25941
 \end{aligned}$$

L'effectif théorique est donc $0,25941 \times 200 = 51,88$

– Pour la classe [9,5;10] :

$$\begin{aligned}
 P(9,5 < X < 10) &= P(X < 10) - P(X < 9,5) \\
 &= P\left(U < \frac{10 - 9,535}{0,6919}\right) - P(U < -0,77) \\
 &= P(U < 0,67) - P(U < -0,05) \\
 &= P(U < 0,67) - 1 + P(U < 0,05) \\
 &= 0,74857 - 1 + 0,51994 = 0,26851
 \end{aligned}$$

L'effectif théorique est donc $0,26851 \times 200 = 53,70$

– Pour la classe [10 ; 10,5] :

$$\begin{aligned} P(10 < X < 10,5) &= P(X < 10,5) - P(X < 10) \\ &= P\left(U < \frac{10,5 - 9,535}{0,6919}\right) - P(U < 0,67) \\ &= P(U < 1,39) - P(U < 0,67) \\ &= 0,91774 - 0,74857 = 0,16917 \end{aligned}$$

L'effectif théorique est donc $0,16917 \times 200 = 33,83$

– Pour la classe [10,5 ; 11] :

$$\begin{aligned} P(10,5 < X < 11) &= P(X < 11) - P(X < 10,5) \\ &= P\left(U < \frac{11 - 9,535}{0,6919}\right) - P(U < 1,39) \\ &= P(U < 2,12) - P(U < 1,39) \\ &= 0,98300 - 0,91774 = 0,06526 \end{aligned}$$

L'effectif théorique est donc $0,06526 \times 200 = 13,05$

– Pour la classe [11 ; 12] il reste :

$$200 - 13,05 - 33,83 - 53,7 - 51,88 - 30,77 - 13,03 = 3,74$$

On choisit donc de faire fusionner cette classe avec la précédente et on a ainsi pour la classe [10,5 ; 12] un effectif théorique de $13,05 + 3,74 = 16,79$

On résume les résultats ainsi avec le calcul du χ^2 de chaque case :

Cours	[7,5 ; 8,5[[8,5 ; 9[[9 ; 9,5[[9,5 ; 10[[10 ; 10,5[[10,5 ; 12[
Nombre de jours	13	20	64	66	22	15
Effectif théorique	13,03	30,77	51,88	53,7	33,83	16,79
χ^2	0,00	3,77	2,83	2,82	4,14	0,19

La somme des χ^2 de chaque case est :

$$0,00 + 3,77 + 2,83 + 2,82 + 4,14 + 0,19 = 13,75$$

Cette valeur est la valeur du χ^2 observé.

Il y a 6 cases dans ce tableau et le nombre de degrés de paramètres de la loi normale est 2. Le nombre de degrés de libertés est donc $6 - 1 - 2 = 3$

D'après la table du χ^2 pour $\alpha = 1\%$, on a une valeur de χ^2 de 11,35

Ainsi χ^2 de la table $< \chi^2$ observé, donc au risque 1 % de rejeter à tort le modèle alors que celui-ci est légitime, on rejette le modèle de la loi normale pour cette série.

Par contre, si on fixe un risque de 0,1 % de rejeter à tort le modèle alors que celui-ci est légitime, la table propose un χ^2 de 16,27 ce qui nous

conduit à la conclusion inverse; autrement dit, pour $\alpha = 0,1 \%$, on valide l'utilisation du modèle.

17. Les actions A et les actions D évoluent en sens inverse d'après l'observation réalisée sur la première semaine. Le calcul de la covariance permet de confirmer l'observation ; la question 3 indique une covariance des deux titres de $-1,15 < 0$.

Ainsi, dans un portefeuille constitué à 80 % d'actions A et à 20 % d'actions D, le mouvement principal sera donné par les actions A qui sont majoritaires. Les actions D évolueront *a priori* en contresens du mouvement du portefeuille. En période de hausse des cours de l'action A, elles vont alourdir la croissance du portefeuille, ce qui va réduire les profits éventuels en cas de revente de l'ensemble des titres. Par contre, en période de baisse du cours des actions A, ces actions D vont remonter et amoindrir les pertes éventuelles. Les actions D – minoritaires dans le portefeuille – vont avoir un effet global de nivellement. Elles réduisent ainsi la prise de risque liée à l'investissement dans les actions A et méritent le terme d'« actions de sécurité ».

18. La valeur globale du portefeuille de cet investisseur dépend de la somme investie sur l'ensemble de la semaine. *A priori* cette somme est de $10\,000 \times 5 = 50\,000$ euros. Cependant, il est important de remarquer que l'intégralité des 10 000 euros ne peut être investie chaque jour.

En effet, chaque jour l'investisseur va acheter pour 8 000 euros d'actions A et 2 000 euros actions B. Ceci ne correspond pas toujours à un nombre entier d'actions. Le jour 1, le cours des actions A est de 10,1. L'investisseur peut donc acheter $\frac{8\,000}{10,1} = 792,08$ actions. Or, il n'est possible d'acheter qu'un nombre entier d'actions. On supposera alors que la somme investie est une limite à ne pas dépasser. Ainsi l'investisseur va acheter seulement 792 actions A le premier jour, ce qui représente un investissement réel de $792 \times 10,1 = 7\,999,2$ euros. De même pour les actions D, le premier jour on a $\frac{2\,000}{15,5} = 129,03$. L'investisseur achète donc 129 actions pour une somme de $129 \times 15,5 = 1\,999,5$ euros.

L'investissement réel le jour 1 est donc de $7\,999,2 + 1\,999,5 = 9\,998,7$ euros.

On résume dans les tableaux suivant les sommes investies et le nombre d'actions achetées.

Pour les actions A :

Jour	Cours	Nombres d'actions	Somme investie
1	10,1	792	7 999,2
2	9,8	816	7 996,8
3	9,7	824	7 992,8
4	10,2	784	7 996,8
5	10,8	740	7 992
Total		3 956	39 970,6

Pour les actions D :

Jour	Cours	Nombres d'actions	Somme investie
1	15,5	129	1 999,5
2	18,3	109	1 994,7
3	20,3	98	1 989,4
4	14,3	139	1 987,7
5	11,2	178	1 993,6
Total		653	9 964,9

Au total la somme investie est donc de :

$$39\,970,6 + 9\,964,9 = 49\,935,5 \text{ euros}$$

19. Le cours moyen se détermine par le rapport de la somme investie sur le nombre d'actions achetées par catégorie.

– Pour les actions A, le cours moyen est de $\frac{39\,970,6}{3\,956} = 10,10$ euros par action.

– Pour les actions D, le cours moyen est de $\frac{9\,964,9}{653} = 15,26$ euros par action.

20. Le cours moyen des actions A était de 10,12 la semaine 1 et le cours des actions D était de 15,92. On remarque dans les deux cas que le cours du portefeuille est inférieur au cours moyen des actions sur la semaine boursière. Cela s'explique par le fait que l'investisseur a acheté plus d'actions les jours où les cours sont plus bas. Les actions moins coûteuses et plus nombreuses ont tiré la moyenne vers le bas.

Avoir un cours de portefeuille inférieur au cours moyen des actions sur la semaine est une bonne chose pour un acheteur qui souhaite voir les cours remonter et revendre vite.

21. On a d'une part la covariance entre A et D = -1,1516 et d'autre part l'écart-type de D = 3,16. Ainsi le coefficient directeur de la droite des moindres carrés est $a = \frac{-1,1516}{3,16^2} = -0,1153$.

La moyenne des cours de l'action A est de 10,12 et celle des cours de l'action D est 15,92. Ainsi l'ordonnée à l'origine de la droite des moindres carrés est $b = 10,12 - 15,92 \times (-0,1153) = 11,9556$.

L'équation de la droite des moindres carrés est donc :
 $y = -0,1153 \times x + 11,9556$

22. On a d'une part la covariance entre A et D = -1,1516 et d'autre part l'écart-type de D = 3,16. Ainsi le coefficient directeur de la droite des moindres carrés est $a = \frac{-1,1516}{3,16^2} = -0,1153$.

La moyenne des cours de l'action A est de 10,12 et celle des cours de l'action D est 15,92. Ainsi l'ordonnée à l'origine de la droite des moindres carrés est $b = 10,12 - 15,92 \times (-0,1153) = 11,9556$.

L'équation de la droite des moindres carrés est donc :

$$y = -0,1153 \times x + 11,9556$$

Le cours de l'action A est de 11,4 le vendredi de la semaine 2. On a donc $y = 11,4$ selon la modélisation par une droite des moindres carrés.

$$\text{Ainsi } 11,4 = -0,1153 \times x + 11,9556$$

$$\text{Soit } 11,4 - 11,9556 = -0,1153 \times x$$

$$\text{D'où } -0,5556 = -0,1153 \times x$$

$$\text{Et } x = 4,8185$$

23. Si l'investisseur vend l'ensemble de ses titres le vendredi de la semaine 2, il va gagner avec les actions A : $3\,956 \times 11,4 = 45\,098,4$ et $653 \times 4,8185 = 3\,146,48$ avec les actions D.

Il aura donc gagné en tout $45\,098,4 + 3\,146,48 = 48\,244,9$ euros.

La somme investie était de 49 935,5 euros. L'investisseur a donc perdu de l'argent. La perte est de $49\,935,5 - 48\,244,9 = 1\,690,62$

24. Le résultat est surprenant. En effet, les actions D semblent perturber le portefeuille de manière importante. Elles perdent leur effet de sécurité pour devenir nocives. Le cours prévu pour les actions D de 4,8185 est surprenant aussi, car il sort de l'intervalle habituel des valeurs de l'action D.

En calculant le coefficient de corrélation linéaire entre les cours de l'action A et ceux de l'action D, on peut estimer la qualité d'une régression linéaire.

On a une covariance de $-1,1516$, l'écart-type de A est de $0,3868$ et l'écart-type de D est de $3,16$. Ainsi on trouve un coefficient de corrélation linéaire $r = \frac{-1,1516}{0,3868 \times 3,16} = -0,94$

Le résultat est proche de -1 . La régression linéaire semble donc une bonne chose. Alors pourquoi un résultat aussi surprenant ?

Le modèle n'est construit que sur 5 observations, ce qui est très peu en finance de marché. Pour établir un tel modèle de régression linéaire, il faudrait considérer des milliers de valeurs. De plus, dans ce modèle, on prend en compte les valeurs réelles des cours. On sait qu'il est plus pertinent de s'intéresser aux variations relatives.